

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:46:24 ; Search time 21.7719 seconds
(without alignments)
2086.045 Million cell updates/sec

Title: US-09-884-319A-2

Perfect score: 962
Sequence: 1 IPKVDLRWQDCEDCDRTTRG.....VNIGACCPVCLQKRAEKP 176

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mhc:*
- 8: sp-organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-virus:*
- 16: sp-bacteriopl:*
- 17: sp-archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	962	100.0	1496	4	Q92626
2	942.5	98.0	221	4	Q96GFS
3	491.5	51.1	224	11	Q99MGO
4	390.5	40.6	494	4	Q96LH9
5	186	19.3	1311	5	Q961K8
6	186	19.3	1527	5	Q9VZ24
7	164	17.0	2327	13	Q91BG7
8	163	16.9	1048	4	Q8AMW5
9	160	16.6	1036	4	Q9NZV1
10	157	16.3	286	11	Q8BS96
11	157	16.3	1028	11	Q9JLLO
12	153.5	16.0	1535	5	Q23991
13	152.5	15.9	1447	13	Q91B91
14	143.5	14.9	1637	6	Q9XSV8
15	143.5	14.9	5146	6	Q8SPM4
16	139.5	14.3	3198	5	Q26639

17	139	14.4	1172	11	Q8CG21	Q8CG21 mus musculus
18	137	14.2	1376	5	Q26637	Q26637 strongyloce
19	134.5	14.0	1823	5	Q26638	Q26638 paracentrot
20	133.5	13.9	940	13	Q57465	Q57465 gallus gall
21	132.5	13.8	1497	11	Q61431	Q61431 mus musculus
22	126.5	13.1	441	4	Q96CW7	Q96CW7 homo sapien
23	125.5	13.0	4998	11	Q8CG65	Q8CG65 mus musculus
24	125	12.5	1171	11	Q8CG82	Q8CG82 mus musculus
25	124.5	12.9	281	11	Q8BNM3	Q8BNM3 mus musculus
26	123.5	12.8	104	4	Q96CBO	Q96CBO homo sapien
27	123.5	12.8	413	4	Q8NZW7	Q8NZW7 homo sapien
28	123.5	12.8	503	4	Q8TEH7	Q8TEH7 homo sapien
29	123.5	12.8	1258	13	Q8AW11	Q8AW11 brachydantio
30	119.5	12.4	210	6	Q8HVS3	Q8HVS3 sus scrofa
31	119.5	12.4	1458	13	Q910B9	Q910B9 oncorhynch
32	115	12.0	784	11	Q8BM43	Q8BM43 mus musculus
33	115	12.0	816	11	Q8R417	Q8R417 mus musculus
34	115	12.0	858	11	Q8BM06	Q8BM06 mus musculus
35	115	12.0	984	5	Q96411	Q96411 halocynthia
36	114	11.9	955	4	Q96DN2	Q96DN2 homo sapien
37	113	11.7	206	13	Q90802	Q90802 gallus gall
38	113	11.7	355	4	Q8N9Z6	Q8N9Z6 homo sapien
39	111.5	11.6	406	11	Q92513	Q92513 mus musculus
40	111.5	11.6	426	11	Q8VEA6	Q8VEA6 mus musculus
41	111	11.5	1491	13	Q91718	Q91718 xenopus lae
42	110.5	11.5	324	11	Q8C8N3	Q8C8N3 mus musculus
43	110	11.4	177	6	Q62801	Q62801 canis famill
44	110	11.4	216	6	Q9G1M4	Q9G1M4 canis famill
45	109	11.3	101	4	Q9Y3H8	Q9Y3H8 homo sapien

ALIGNMENTS

RESULT 1

Q92626 PRELIMINARY; PRT; 1496 AA.

AC Q92626; 01-FEB-1997 (TREMBLrel. 02, Created)

DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE MBL0BLAST KIAA0230 (Fragment).

GN KIAA0230.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RX MEDLINE=97191544; PubMed=9039502;

RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;

RT Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain.;"

RL DNA Res. 3:321-329(1996).

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Melanoma;

RX MEDLINE=95048383; PubMed=7959781;

RA Trent J.M., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,

RT "Assignment of a human melanoma associated gene MG50 (D2S448) to chromosome 2p25.3 by fluorescence in situ hybridization.;"

RL Genomics 22:243-244(1994).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Melanoma;

RA Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;

RT "Identification of a novel melanoma gene (MG50) - likely the gene for IL-1 receptor antagonist - which encodes epitopes recognized by human cytolytic T lymphocytes.;"

```

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; D86983; BAAL3219.1; -.
DR EMBL; AF200348; AAF06354.1; -.
DR HSSP; P05164; ICXP.
DR InterPro; IPR002007; Anim_peroxidase.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_Typ.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF03098; An_peroxidase; 1.
DR Pfam; PF00047; Ig; 4.
DR Pfam; PF00560; LRR; 5.
DR Pfam; PF01463; LRCT; 1.
DR Pfam; PF00093; VWC; 1.
DR PRINTS; PR00457; ANPEROXIDASE.
DR SMART; SM00408; IGC2; 4.
DR SMART; SM00082; LRRT; 1.
DR SMART; SM00013; LRNT; 1.
DR SMART; SM00369; LRR_Type; 4.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS30835; IG_Like; 4.
DR PROSITE; PS50506; LRR_TYPICAL; 1.
DR PROSITE; PS01208; VWF_C; 1.
KW Immunoglobulin domain.
FT NON_TER.
SQ SEQUENCE 1496 AA; 167209 MW; E9B9A7069BFIAPF CRC64;

Query Match
Best Local Similarity 100.0%; Score 962; DB 4; Length 1496;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IPRVDLRWQDCCEDCRRGQFNAFSYHFRGRSLSEFSYQEDKPTKTRPKRIPSVGRG 60
DB 1321 IPRVDLRWQDCCEDCRRGQFNAFSYHFRGRSLSEFSYQEDKPTKTRPKRIPSVGRG 1380
OY 61 EHLNSTSAFSTRDASGTNDFREFVLEMOKTTIDLRQIKKLSRLSTTECVADAGGSH 120
DB 1381 EHLNSTSAFSTRDASGTNDFREFVLEMOKTTIDLRQIKKLSRLSTTECVADAGGSH 1440
OY 121 ANNTKMKKDACCTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCIQKRAEEXP 176
DB 1441 ANNTKMKKDACCTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCIQKRAEEXP 1496

RESULT 2
O96GF5 PRELIMINARY; PRT; 221 AA.
AC O96GF5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR InterPro; IPR002007; Anim_peroxidase.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF03098; An_peroxidase; 1.
DR Pfam; PF00093; VWC; 1.
DR SMART; SM00214; VWC; 1.
KW Hypothetical protein.
FT NON_TER.

```

```

SQ SEQUENCE 221 AA; 24819 MW; 253E54F76B126E8E CRC64;

Query Match
Best Local Similarity 98.0%; Score 942.5; DB 4; Length 221;
Matches 175; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1 IPRVDLRWQDCCEDCRRGQFNAFSYHFRGRSLSEFSYQEDKPTKTRPKRIPSVGRG 60
DB 47 IPRVDLRWQDCCEDCRRGQFNAFSYHFRGRSLSEFSYQEDKPTKTRPKRIPSVGRG 106
OY 61 EHLNSTSAFSTRDASGTNDFREFVLEMOKTTIDLRQIKKLSRLSTTECVADAGGSH 120
DB 107 EHLNSTSAFSTRDASGTNDFREFVLEMOKTTIDLRQIKKLSRLSTTECVADAGGSH 166
OY 121 ANNTKMKKDACCTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCIQKRAEEXP 176
DB 167 ANNTKMKKDACCTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCIQKRAEEXP 221

RESULT 3
O99MG0 PRELIMINARY; PRT; 224 AA.
AC O99MG0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Melanoma-associated antigen MG50 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA Long X., Bigsby R.M., Nephew K.P.;
RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF346790; AAK30128.1; -.
DR InterPro; IPR002007; Anim_peroxidase.
DR Pfam; PF03098; An_peroxidase; 1.
FT NON_TER.
SQ SEQUENCE 224 AA; 25609 MW; D3EF84B599B192A5 CRC64;

Query Match
Best Local Similarity 51.1%; Score 491.5; DB 11; Length 224;
Matches 102; Conservative 12; Mismatches 33; Indels 31; Gaps 5;

OY 1 IPRVDLRWQDCCEDCRRGQFNAFSYHFRGRSLSEFSYQEDKPTKTRPKRIPSVGRG 60
DB 48 IPRVDLRWQDCCEDCRRGQFNAFSYHFRGRSLSEFSYQEDKPTKTRPKRIPSVGRG 106
OY 61 EHLNSTSAFSTRDASGTNDFREFVLEMOKTTIDLRQIKKLSRLSTTECVADAGGSH 120
DB 107 EHLNSTSAFSTRDASGTNDFREFVLEMOKTTIDLRQIKKLSRLSTTECVADAGGSH 155
OY 121 ANNTKMKKDACCTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCIQKRAEEXP 164
DB 156 -QSVWMTVIVNPRQOHEVGRKROMPYCECKNGQVTCFEVACHLQPA--VKVESACC 208

RESULT 4
O96LH9 PRELIMINARY; PRT; 494 AA.
AC O96LH9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ25471.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

```


RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE-Testis:
 RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
 RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
 RA Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
 RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Katsuka R.,
 RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terasima Y.,
 RA Matanabe M., Suzuki Y., Hata H., Nakagawa K., Mizuno S., Morinaga M.,
 RA Kawamura M., Sugiyama T., Itie R., Otsuki T., Sato H., Nishikawa T.,
 RA Sugiyama A., Kawakami B., Nagai K., Isogai T., Sugano S.,
 RT "NEO human cDNA sequencing project."
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK058200; BAB71713.1; -
 DR InterPro: IPR002007; Anim_peroxidase;
 DR InterPro: IPR003328; Tila_Cysrich.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF03098; An_peroxidase; 1.
 DR Pfam: PF02345; Tila; 1.
 DR Pfam: PF00093; vwc; 1.
 DR PRINTS: PR00457; ANPEROXIDASE.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWF; 1.
 DR PROSITE: PS01208; VWF; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 494 AA; 55891 MW; ACE7AAE882E08519 CRC64;

Query Match 40.6%; Score 390.5; DB 4; Length 494;
 Best Local Similarity 45.7%; Pred. No. 2.6e-29;
 Matches 79; Conservative 18; Mismatches 61; Indels 15; Gaps 3;

QY 1 IPRVLDLWVQDCCEDCRTRGQFNAPSYHFRGRSLSEFYOEKDPKTRPRKIPSVGRQ- 59
 DB 318 IPRVLDLWVQDCCEDCRTRGQFNAPSYHFRGRSLSEFYOEKDPKTRPRKIPSVGRQ- 59
 QY 60 -----GEHLSNSTSAFSTRSDASGNTDFEFLVEMOKITDTRQIKTESRLSTECVD 114
 DB 373 DKIVYGEDARNTYVAKTKF-----SODFSTFAAEIOTETITALREQINKLEARIROAGCD 428
 QY 115 AGGESHANTTKWKKDKACTICECKDGVGVTFEACPPATCAVPINIGACCPVC 167
 DB 429 VAGVPRKAEERWKEDCTHCICESGQVTCVETICPPAPPSPELVAGTCCPVC 481

RESULT 5
 ID 0961K8 PRELIMINARY; PRT; 1311 AA.
 AC 0961K8;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE GH18946P.
 GN PXN OR CG12002.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA Stapleton N., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AT051536; AAK92960.1; -
 DR FlyBase; Fgn0011828; Pxn.
 DR InterPro: IPR002007; Anim_peroxidase.
 DR InterPro: IPR007110; Ig-Like.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR001007; VWF_C.

DR Pfam: PF03098; An_peroxidase; 1.
 DR Pfam: PF00047; Ig; 4.
 DR Pfam: PF00093; vwc; 1.
 DR PRINTS: PR00457; ANPEROXIDASE.
 DR SMART: SM00408; IGC2; 3.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS05835; IG_LIKE; 4.
 DR PROSITE: PS01208; VWF; 1.
 KW Immunoglobulin domain.
 SQ SEQUENCE 1311 AA; 145961 MW; 4416DFEE3A2912A CRC64;

Query Match 19.3%; Score 186; DB 5; Length 1311;
 Best Local Similarity 25.9%; Pred. No. 3.8e-09;
 Matches 51; Conservative 27; Mismatches 71; Indels 48; Gaps 6;

QY 4 VDLRWQDCCEDCRTRGQFNAPSYHFRGRSLSEFYOEKDPKTRPRKIPSVGRGHL 63
 DB 1126 INLYWQEC-----GRCSPPALFD-----SYIPQYTKRSNRQ-RDLGKNDENV 1170
 QY 64 SNSTSAFSTRSDASGNTDFR-----EVLDMOKITDTRQIKLE----- 104
 DB 1171 ATAESYDPLESLIDYNEERVSGLLELISFQREKLKHLKLELSDSCNSADSEPVAV 1230
 QY 105 -----SRLSTTECVADAGGESHANTTKWKKDKACTICECKDGVGVTFEACPPATC 153
 DB 1231 VOLAAPQLVSKPRKRSKHCVDKGTTRLNNEMWSPVVCIKCNCFHQVNCLEHRCGEVSC 1290
 QY 154 ---AVNVPNGACCPVC 167
 DB 1291 PCGVDPITPEPEACCPHC 1307

RESULT 6
 ID 09VZ24 PRELIMINARY; PRT; 1527 AA.
 AC 09VZ24;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE PXN protein.
 GN PXN OR CG12002.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Frannkooch C., Baldwin D.,
 RA Ballou R.M., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwu C.,
 RA Jalili M., Kalush F., Kappen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,


```

Q23991
ID Q23991 PRELIMINARY; PRT: 1535 AA.
AC Q23991;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE PEROXIDASE precursor.
GN PXN OR CG12002.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
OC NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RX MEDLINE=9434125; PubMed=8062820;
RA Nelson R.E., Fessler L.I., Takagi Y., Blumberg B., Keene D.R.,
RA Olson P.F., Parker C.G., Fessler J.H.;
RT "Peroxidase: a novel enzyme-matrix protein of Drosophila
RT development.";
RL EMBL: U11052; AAA61568.1; -
DR HSSP: P05164; 1CXP.
DR Flyase; FBgn0011828; PXN.
DR InterPro: IPR002007; Anim_peroxidase.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003598; Ig_C2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003591; LRR_Typ.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF03098; An_peroxidase; 1.
DR Pfam: PF00047; Ig_4.
DR Pfam: PF00560; LRR_5.
DR Pfam: PF01462; LRRNT; 1.
DR PRINTS: PR00457; ANPEROXIDASE.
DR SMART: SM00408; IGC2; 3.
DR SMART: SM00013; LRRNT; 1.
DR SMART: SM00369; LRR_Typ; 2.
DR SMART: SM00214; VWC; 1.
DR PROSITE: PS50835; IG_Like; 4.
DR PROSITE: PS50506; LRR_TYPICAL; 1.
DR KIMMUNOGLOBULIN domain; Signal.
FT SIGNAL 1 23
FT CHAIN 24 1535
FT SEQUENCE 1535 AA; 171084 MW; 1B2D7A0C76339D64 CRC64;

Query Match 16.0%; Score 153.5; DB 5; Length 1535;
Best Local Similarity 23.6%; Pred. No. 6.1e-06;
Matches 47; Conservative 27; Mismatches 68; Indels 57; Gaps 7;

QY 4 VDLRWQDCCECDECRGQFNA---FSYHRCGRSLFEFVQEDCKPTKTRPRKIPSVGRQ 59
DB 1345 INILYMQEC-----GRCSPPAIVPYR-----KPI-PSGDRKQKRLDGRK 1384
QY 60 GEHLNSTSAFSTRDASGTNDFR---EVLQMKITTDLTQTKLE----- 104
DB 1385 NEVATAESYDPSLELYDVNERVSGLELIGSFQKELKHLKRLKLEDCNSADSEF 1444
QY 105 -----SRLSTTECVDAAGGSHANNNTWKKDACITCECKDGQVTCFVEACP 149
DB 1445 VQVQVLAAPQVLSKPRKSHCVDDKGTTRLNNEVMSPDVCTKCNCFHGQVNCLEKRG 1504
QY 150 PATC---AVPVNIPGACCP 165
DB 1505 EVSCPPGVPLTASGGLP 1523

```

```

AC Q91B91;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Type I collagen alpha 1.
GN COL1A1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_TaxId=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Goto T., Katada T., Kinoshita T., Kubota H.Y.;
RT "Expression and characterization of Xenopus type I collagen alpha 1
RT (COL1A1) during embryonic development.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB034701; BA94972.1; -
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR Pfam: PF00093; VWC; 1.
DR ProDom: PD000007; Collagen; 2.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWC; 1.
DR PROSITE: PS01208; VWF_C; 1.
KW Collagen.
SQ SEQUENCE 1447 AA; 137446 MW; AAA6DD24158E3B8 CRC64;

Query Match 15.9%; Score 152.5; DB 13; Length 1447;
Best Local Similarity 38.6%; Pred. No. 7.2e-06;
Matches 27; Conservative 11; Mismatches 31; Indels 1; Gaps 1;

QY 104 ESRSTTECVDAAGGSHANNNTWKKDACITCECKDGQVTCFVEACPATCAVPVNIAGC 163
DB 25 EHDVOTSDCVQ-HGITYSNRDVWKPEDCOICVDMNGNIIIDBVMGEDADCPNPVITVPGEC 83
QY 164 CPVCLQKRAE 173
DB 84 CPVCTDNDQAQ 93

RESULT 14
Q9XSV8 PRELIMINARY; PRT: 1637 AA.
ID Q9XSV8;
AC Q9XSV8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE SCO-spondin (Fragment).
GN SCO-SPONDIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxId=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Gobron S., Creveaux I., Didier R., Meinzel R.;
RT "Characterization of cattle SCO-spondin.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: A1133488; CAB46239.1; -
DR HSSP: P56682; 1CXY.
DR InterPro: IPR006206; Cys_knot.
DR InterPro: IPR006207; Cys_knot_C.
DR InterPro: IPR001545; Gly_hormoneB.
DR InterPro: IPR002919; TIL_Cysrich.
DR InterPro: IPR000884; TSPI.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF00007; Cys_knot; 1.

```

DR Pfam: PF01826; TIL; 3.
 DR Pfam: PF00090; tsp_1; 11.
 DR SMART: SM00041; CT; 1.
 DR SMART: SM00068; GHB; 1.
 DR SMART: SM00209; TSP1; 11.
 DR SMART: SM00214; VWC; 2.
 DR PROSITE: PS01225; CTCK_2; 1.
 DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; 2.
 DR PROSITE: PS50092; TSP1; 11.
 DR PROSITE: PS01208; VWF; 1.
 FT NON_TER 1 1
 SQ SEQUENCE 1637 AA; 173347 MW; 4C5BAD1DD346C925 CRC64;
 Query Match 14.9%; Score 143.5; DB 6; Length 1637;
 Best Local Similarity 42.9%; Pred. No. 6.1e-05;
 Matches 27; Conservative 7; Mismatches 24; Indels 5; Gaps 3;
 QY 117 GESHANTKWKKDACITCECKDGOVTCFVACPATCA---VPVNIAGACPVCLQRAE 173
 Db 1480 GRHPGSEWOK-ACESCRCVSGESIC-TQHCPLTCAOGETAVQEPGGCCPTCRQEAPE 1537
 QY 174 EKP 176
 Db 1538 EOP 1540

RESULT 15

OSSPM4 PRELIMINARY; PRT; 5146 AA.
 AC OSSPM4;
 DT 01-JUN-2002 (TREMBREL. 21, Created)
 DT 01-JUN-2002 (TREMBREL. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBREL. 23, Last annotation update)
 DE SCO-spondin.
 GN SCO-SPONDIN.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Subcommisural organ;
 RA Meinel A.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Subcommisural organ;
 RX MEDLINE=20465125; PubMed=11008217;
 RA Gobron S.;
 RT "Subcommisural organ/Reissner's fiber complex: characterization of
 RT SCO-spondin, a glycoprotein with potent activity on neurite
 RT outgrowth.";
 RL Gila 32:177-191(2000).
 DR EMBL: AJ16457; CAC94914.1; -;
 DR InterPro: IPR001064; Crystallin.
 DR InterPro: IPR006208; Cys_knot.
 DR InterPro: IPR006207; Cys_knot_C.
 DR InterPro: IPR000421; FA58_C.
 DR InterPro: IPR001545; Gly_hormoneB.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR002919; TIL_Cysrich.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR001007; VWF_C.
 DR InterPro: IPR001846; VWF_D.
 DR Pfam: PF00007; Cys_knot; 1.
 DR Pfam: PF00754; F5_F8_Type_C; 1.
 DR Pfam: PF00057; Id1_recept_A; 10.
 DR Pfam: PF01826; TIL; 10.
 DR Pfam: PF00090; tsp_1; 25.
 DR Pfam: PF00094; vwd; 3.
 DR SMART: SM00041; CT; 1.
 DR SMART: SM00231; FA58C; 1.

DR SMART: SM00068; GHB; 1.
 DR SMART: SM00192; LDla; 10.
 DR SMART: SM00209; TSP1; 25.
 DR SMART: SM00216; VWD; 3.
 DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 1.
 DR PROSITE: PS01225; CTCK_2; 1.
 DR PROSITE: PS01285; FA58C_1; 1.
 DR PROSITE: PS01286; FA58C_2; 1.
 DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; 2.
 DR PROSITE: PS01209; LDLRA_1; 7.
 DR PROSITE: PS50068; LDLRA_2; 9.
 DR PROSITE: PS50092; TSP1; 25.
 DR PROSITE: PS01208; VWF; 1.
 SQ SEQUENCE 5146 AA; 543576 MW; 724C5FB8727E13DA CRC64;
 Query Match 14.9%; Score 143.5; DB 6; Length 5146;
 Best Local Similarity 42.9%; Pred. No. 0.00022;
 Matches 27; Conservative 7; Mismatches 24; Indels 5; Gaps 3;
 QY 117 GESHANTKWKKDACITCECKDGOVTCFVACPATCA---VPVNIAGACPVCLQRAE 173
 Db 4989 GRHPGSEWOK-ACESCRCVSGESIC-TQHCPLTCAOGETAVQEPGGCCPTCRQEAPE 5046
 QY 174 EKP 176
 Db 5047 EOP 5049

Search completed: July 24, 2003, 12:56:48
 Job time : 24.7719 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:28:44 ; Search time 4.86524 Seconds

(without alignments)
1701.191 Million cell updates/sec

Title: US-09-884-319A-2

Sequence: 1 IPRVDLRVWQDCEDCERTRG.....VNIPGACPVCIQRAEEKP 176

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	148	15.4	1178	1 TSP2_CHICK
2	147	15.3	941	1 CHRD_XENLA
3	147	15.3	1172	1 TSP2_HUMAN
4	143	14.9	1170	1 TSP2_BOVIN
5	139	14.4	1172	1 TSP2_MOUSE
6	137.5	14.3	810	1 NEIL_HUMAN
7	134.5	14.0	810	1 NEIL_RAT
8	129	13.4	450	1 NRUL_HUMAN
9	129	13.4	456	1 NRUL_CHICK
10	127	13.2	699	1 ECM2_HUMAN
11	125	13.0	447	1 NRUL_MOUSE
12	125	13.0	1170	1 TSP1_HUMAN
13	125	13.0	1170	1 TSP1_MOUSE
14	123.5	12.8	955	1 CHRD_HUMAN
15	123.5	12.8	1496	1 CA25_HUMAN
16	117	12.2	1170	1 TSP1_BOVIN
17	114	11.9	846	1 NEIL_MOUSE
18	113	11.7	940	1 CHRD_MOUSE
19	113	11.7	1173	1 TSP1_XENLA
20	112.5	11.7	940	1 CHRD_BRARE
21	110	11.4	1460	1 CALI_CANFA
22	108.5	11.3	114	1 MSP2_SAGOE
23	108	11.2	1466	1 CAL3_HUMAN
24	107	11.1	1453	1 CAL1_CHICK
25	107	11.1	1464	1 CAL1_HUMAN
26	106.5	11.1	816	1 NEIL_RAT
27	106	11.0	816	1 NEIL_CHICK
28	106	11.0	1038	1 SEG_DROME
29	105.5	11.0	1459	1 CAL2_MOUSE
30	104.5	10.9	176	1 CHRD_RAT
31	104	10.8	113	1 MSMB_MOUSE
32	104	10.8	1262	1 CAL3_CHICK
33	103.5	10.8	816	1 NEIL2_HUMAN

34	102.5	10.7	3133	1 HMCT_BOMO
35	102	10.6	1213	1 JAG3_BRARE
36	102	10.6	1464	1 CAL3_MOUSE
37	100	10.4	112	1 MSPA_SAGOE
38	97	10.1	114	1 MSMB_MACMU
39	97	10.1	114	1 MSMB_PAPAN
40	96.5	10.0	113	1 MSMB_RAT
41	92	9.6	114	1 MSMB_HUMAN
42	89	9.3	111	1 MSMB_PIG
43	89	9.3	114	1 MSPE_SAGOE
44	87	9.0	632	1 FMN2_HUMAN
45	87	9.0	5376	1 ZAN_MOUSE

ALIGNMENTS

RESULT 1	TSP2_CHICK	STANDARD;	PRT; 1178 AA.
ID	TSP2_CHICK		
AC	P35440;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Thrombospondin 2 precursor.		
GN	THBS2 OR TSP2.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91217026; PubMed=2022631;		
RA	Lawler J., Duquette M., Ferro P.;		
RT	"Cloning and sequencing of chicken thrombospondin.;"		
RL	J. Biol. Chem. 266:8039-8043(1991).		
CC	- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.		
CC	- SUBUNIT: Homotrimer; disulfide-linked.		
CC	- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.		
CC	- SIMILARITY: Contains 1 WFEC domain.		
CC	- SIMILARITY: Contains 3 EGF-like domains.		
CC	- SIMILARITY: Contains 3 TSP type-1 domains.		
CC	- SIMILARITY: Contains 7 TSP type-3 domains.		
CC	- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M60853; AAA51437.1; -		P98092 bombyx mori
DR	PIR; A39804; A39804.		O9054 brachydanio
DR	HSP; P00740; 1EDM.		P08121 mus musculus
DR	InterPro; IPR001881; EGF_Ca.		O97936 saguinus oe
DR	InterPro; IPR006209; EGF_like.		P25142 macaca mula
DR	InterPro; IPR006210; IEGF.		O28767 papio anubi
DR	InterPro; IPR000884; TSP1.		P97580 rattus norv
DR	InterPro; IPR003367; TSPN.		P08118 homo sapien
DR	InterPro; IPR003129; TSPN.		O02826 sus scrofa
DR	InterPro; IPR001007; WFEC_C.		O97935 saguinus oe
DR	Pfam; PF00008; EGF; 2.		O9n256 homo sapien
DR	Pfam; PF00090; TSP_1; 3.		O88799 mus musculu
DR	Pfam; PF02412; TSP_3; 8.		
DR	Pfam; PF02210; TSPN; 1.		
DR	Pfam; PF00093; VWC; 1.		
DR	SMART; SM00181; EGF; 2.		
DR	SMART; SM00209; TSP1; 3.		

[illegible]

```

OY 78 GTNDFRE-----VLEMQTINDLRFOUKLESRLST-----ECVDAG----- 116
DB 264 KTEKKAEPFCBSCGELGTMFEITGLRVANNLADNLQKXSENOIMMELIGPKTKLTQ 323
OY 117 -----GESHANNTKMKKDACTICECKDQVTCVEACPPATCAVPNIPGACCPVC 167
DB 324 SVCMDGRVAFDSBSWIYDSCTKCTCQDSKRTVCHQITCPVSCADPSFIEGECPCVC 380

RESULT 2
CHRD_XENLA STANDARD; PRT; 941 AA.
ID CHRD_XENLA
AC 091713;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chordin precursor (organizer-specific secreted dorslizing factor).
OS Xenopus laevis (African clawed frog).
OC Euryarchaeta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Dorsal 1lp;
RX MEDLINE=95094250; PubMed=8001117;
RA Sasai Y., Lu B., Steinbeisser H., Gelsert D., Gont L.K.,
RA de Robertis E.M.;
RA "Xenopus chordin: a novel dorslizing factor activated by organizer-
RT specific homeobox genes.";
RL Cell 79:779-790(1994).
CC -1- FUNCTION: POTENT DORSALIZING FACTOR. HAS POTENT AXIS-FORMING
CC ACTIVITIES INCLUDING THE ABILITY TO RECRUIT NEIGHBORING CELLS INTO
CC SECONDARY AXES. REGULATES CELL-CELL INTERACTIONS IN THE ORGANIZING
CC CENTERS OF HEAD, TRUNK AND TAIL DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PRECORDAL PLATE, NOTOCHORD, CHORDONEURAL
CC HINGE; EXPRESSION CLOSELY FOLLOWS THE AREAS OF THE EMBRYO THAT
CC EXPRESS GOOSECOID AND XNOT2.
CC -1- DEVELOPMENTAL STAGE: EMBRYOGENESIS.
CC -1- SIMILARITY: BELONGS TO THE CHORDIN FAMILY.
CC -1- SIMILARITY: Contains 4 WFC domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, L35764; AAC42222.1; -.
CC PIR, A55195; A55195.
CC InterPro, IPR006559; SOG.
CC InterPro, IPR001007; WFC_C.
CC Pfam, PF00093; WFC; 4.
CC SMART, SM00566; SOG; 3.
CC SMART, SM00214; WFC; 4.
CC DR PROSITE, PS01208; WFC_1; 2.
CC DR PROSITE, PS50184; WFC_2; 3.
CC Developmental protein; Repeat; Glycoprotein; signal.
KW SIGNAL
FT CHAIN 1 19 CHORDIN.
FT CHAIN 20 941 CHORDIN.
FT DOMAIN 41 117 WFC 1.
FT DOMAIN 691 751 WFC 2.
FT DOMAIN 769 838 WFC 3.
FT DOMAIN 857 919 WFC 4.
FT CARBOHYD 244 244 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 350 350 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CARBOHYD 735 735 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT SEQUENCE 941 AA; 104947 MW; 7D1C3262064C7333 CRC64;

```



```

Query Match Similarity      15.3%; Score 147; DB 1; Length 941;
Best Local Similarity      29.5%; Pred. No. 3.8e-05;
Matches      43; Conservative      20; Mismatches      45; Indels      38; Gaps      8;

Oy      47      KTRPRKIPSVGRGE-----HLNSTSAFSTRSDASGTDNREFVLEW-----QKTTIDL      96
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db      639      KLNLR-----GEIRQIHIPNSCES-----GGVSLTPPEPEYVEIYEGRGDRDPDL      686

Oy      97      RTQIKLESRLSTECVDAGESHANNTKMKD---ACTICECKDGQYTCVEACPRATC      153
      |      |      |      |      |      |      |      |      |      |      |
Db      687      RK-----DPRACSEE-----GQLRHGSRMAPDYDRKCSVCOKRTVICDPICPLUNC      736

Oy      154      AVPINIPACCPVCLQK---RAEKP      176
      ||      ||      ||      ||      ||      ||      ||      ||      ||      ||
Db      737      SQPVHLDPQCCPVCCEKKMEYVKP      762

RESULT 3
TSP2_HUMAN
ID      TSP2_HUMAN      STANDARD;      PRT:      1172 AA.
AC      P35442.1
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Thrombospondin 2 precursor.
GN      THBS2 OR TSP2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      MEDLINE=94010892; PubMed=8406456;
RX      Label1 T.L., Byers P.H.;
RA      "Sequence and characterization of the complete human thrombospondin 2
RT      cDNA: potential regulatory role for the 3' untranslated region.";
RL      Genomics 17:225-229(1993).
RN      [2]
RP      SEQUENCE OF 560-1172 FROM N.A.
RX      TISSUE=Fibroblast;
RA      MEDLINE=92217961; PubMed=1559694;
RT      Label1 T.L., McGockey Milewicz D.J., Distèche C.M., Byers P.H.;
RT      "Thrombospondin II: partial cDNA sequence, chromosome location, and
RT      expression of a second member of the thrombospondin gene family in
RT      humans.";
RL      Genomics 12:421-429(1992).
RN      [3]
RP      THROMBOSPONDIN REPEATS DISULFIDE BONDS.
RX      MEDLINE=21588233; PubMed=11590138;
RA      Meiselman T.M., Behr A.J., Harris A.C., Annis D.S., Mosher D.F.;
RT      "Disulfide connectivity of recombinant C-terminal region of human
RT      thrombospondin 2.";
RL      J. Biol. Chem. 276:45882-45887(2001).
CC      -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC      CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC      LAMININ AND TYPE V COLLAGEN.
CC      -1- SUBUNIT: Homotrimer; disulfide-linked.
CC      -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC      -1- SIMILARITY: Contains 1 WFDC domain.
CC      -1- SIMILARITY: Contains 3 EGF-like domains.
CC      -1- SIMILARITY: Contains 3 TSP type-1 domains.
CC      -1- SIMILARITY: Contains 7 TSP type-3 domains.
CC      -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).

```

DR	EMBL: L12350; AAA03703.1; -	
DR	EMBL: M81339; -; NOT_ANNOTATED_CDS.	
DR	PIR: A47379; TSHUP2.	
DR	HSSP; P00740; 1EDM.	
DR	Genew; HGNC:11786; THBS2.	
DR	MIM; 188061; -	
DR	GO; GO:0080201; F:heparin binding activity; TAS.	
DR	InterPro; IPR001881; EGF_Ca.	
DR	InterPro; IPR006209; EGF_Like.	
DR	InterPro; IPR006210; IEGF.	
DR	InterPro; IPR000884; TSP1.	
DR	InterPro; IPR003367; tsp_3.	
DR	InterPro; IPR003129; TSPN.	
DR	InterPro; IPR001007; WVF_C.	
DR	Pfam; PF00008; EGF; 2.	
DR	Pfam; PF00090; tsp_1; 3.	
DR	Pfam; PF02412; tsp_3; 9.	
DR	Pfam; PF02210; TSPN; 1.	
DR	Pfam; PF00093; WVC; 1.	
DR	SMART; SM00181; EGF; 3.	
DR	SMART; SM00209; TSP1; 3.	
DR	SMART; SM00210; TSPN; 1.	
DR	SMART; SM00214; WVC; 1.	
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.	
DR	PROSITE; PS01186; EGF_2; 1.	
DR	PROSITE; PS50092; TSP1; 3.	
DR	PROSITE; PS01208; WVF_C_1; 1.	
DR	PROSITE; PS50184; WVC_2; 1.	
KW	Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat	
KW	EGF-like domain; Signal.	
FT	SIGNAL	1
FT	CHAIN	18
FT	DOMAIN	19 1172
FT	DOMAIN	19 215
FT	DOMAIN	19 232
FT	DOMAIN	318 375
FT	DOMAIN	381 431
FT	DOMAIN	437 492
FT	DOMAIN	494 549
FT	DOMAIN	549 589
FT	DOMAIN	590 647
FT	DOMAIN	648 692
FT	DOMAIN	725 760
FT	DOMAIN	761 783
FT	DOMAIN	784 819
FT	DOMAIN	820 842
FT	DOMAIN	843 880
FT	DOMAIN	881 916
FT	DOMAIN	917 952
FT	DOMAIN	953 1172
FT	SITE	928 930
FT	DISULFID	266 266
FT	DISULFID	270 270
FT	DISULFID	393 425
FT	DISULFID	397 430
FT	DISULFID	408 415
FT	DISULFID	449 486
FT	DISULFID	453 491
FT	DISULFID	464 476
FT	DISULFID	506 543
FT	DISULFID	510 548
FT	DISULFID	521 533
FT	DISULFID	553 564
FT	DISULFID	558 574
FT	DISULFID	577 588
FT	DISULFID	594 610
FT	DISULFID	601 619
FT	DISULFID	622 646
FT	DISULFID	652 665
FT	DISULFID	659 678
FT	DISULFID	680 691
FT	DISULFID	707 715
FT	DISULFID	720 740
FT	DISULFID	756 776

```

FT DISULFID 779 799
FT DISULFID 815 835
FT DISULFID 838 858
FT DISULFID 876 896
FT DISULFID 912 932
FT DISULFID 948 1169
FT CARBOHYD 151 151
FT CARBOHYD 316 316
FT CARBOHYD 330 330
FT CARBOHYD 457 457
FT CARBOHYD 584 584
FT CARBOHYD 710 710
FT CARBOHYD 1069 1069
SQ SEQUENCE 1172 AA: 129955 MW: 2AC/BB230E44C6F5 CRC64:

Query Match 15.3% Score 147: DB 1: Length 1172;
Best Local Similarity 48.1% Pred. No. 4.8e-05;
Matches 25; Conservative 3; Mismatches 24; Indels 0; Gaps 0;

QY 117 GESHANNTKWKADCTCECKDGQYTCFVEACPPATCAVPVNI PGACCPVCL 168
DB 324 GRFFAENETWVVDSCYTCCKFKFTICHQITCPATCASPSPFVEGECPSCL 375

RESULT 4
TSP2_BOVIN STANDARD; PRT; 1170 AA.
ID TSP2_BOVIN
AC 095116; Q28180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thrombospondin 2 precursor (Corticotropin-induced secreted protein)
DE (CISP).
GN THBS2 OR TSP2 OR TSP-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Danik M., Chinn A., Lafeuillade M., Keramidas M., Agnesse-Germon S.,
RA Penhoat A., Chen H., Mosher D., Chambaz E.M., Feige J.O.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-522 FROM N.A.
RA MEDLINE=96331130; PubMed=8698834;
RX Lafeuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.O.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells.";
RL J. Cell. Physiol. 167:164-172(1996).
RN [3]
RP SEQUENCE OF 318-831 FROM N.A.
RC TISSUE=Arctic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT TGF-beta.";
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -!- SUBUNIT: Homotrimer; disulfide-linked.
CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -!- SIMILARITY: Contains 1 VWFC domain.
CC -!- SIMILARITY: Contains 3 EGF-like domains.
CC -!- SIMILARITY: Contains 3 TSP type-1 domains.
CC -!- SIMILARITY: Contains 7 TSP type-3 domains.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X96540; CAA65385.1; -.
CC EMBL: X87620; CAA60952.1; -.
CC HSSP: P00740; 1EDM.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR006209; EGF_Like.
CC InterPro: IPR006210; IEGF.
CC InterPro: IPR008844; TSP1.
CC InterPro: IPR003367; TSP_3.
CC InterPro: IPR003129; TSPN.
CC InterPro: IPR001007; VWFC_C.
CC Pfam: PF000090; TSP_1; 3.
CC Pfam: PF02412; TSP_3; 9.
CC Pfam: PF02210; TSPN; 1.
CC Pfam: PF00093; VWC; 1.
CC SMART: SM00181; EGF; 3.
CC SMART: SM00209; TSP1; 3.
CC SMART: SM00210; TSPN; 1.
CC SMART: SM00214; VWC; 1.
CC PROSITE: PS00022; EGF_1; FALSE_NEG.
CC PROSITE: PS01186; EGF_2; 1.
CC PROSITE: PS50092; TSP1; 3.
CC PROSITE: PS01208; VWFC_1; 1.
CC PROSITE: PS50184; VWFC_2; 1.
CC Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
CC EGF-like domain; signal.
FT SIGNAL 1 18
FT CHAIN 1 1170
FT DOMAIN 19 215
FT DOMAIN 19 232
FT DOMAIN 318 375
FT DOMAIN 379 429
FT DOMAIN 435 490
FT DOMAIN 492 547
FT DOMAIN 547 587
FT DOMAIN 588 645
FT DOMAIN 646 680
FT DOMAIN 723 758
FT DOMAIN 759 781
FT DOMAIN 782 817
FT DOMAIN 818 840
FT DOMAIN 841 878
FT DOMAIN 879 914
FT DOMAIN 915 950
FT DOMAIN 951 1170
FT SITE 926 928
FT DISULFID 266 266
FT DISULFID 270 270
FT DISULFID 391 423
FT DISULFID 391 428
FT DISULFID 395 433
FT DISULFID 406 413
FT DISULFID 447 484
FT DISULFID 451 489
FT DISULFID 462 474
FT DISULFID 504 541
FT DISULFID 504 541
FT DISULFID 508 546
FT DISULFID 519 531
FT DISULFID 551 562
FT DISULFID 556 572
FT DISULFID 575 586
FT DISULFID 592 608
FT DISULFID 599 617
FT DISULFID 620 644
FT DISULFID 650 663
FT DISULFID 657 676
FT DISULFID 678 689
FT DISULFID 705 713

```

FT	DISULFID	718	738	BY SIMILIARITY.
FT	DISULFID	754	774	BY SIMILIARITY.
FT	DISULFID	777	797	BY SIMILIARITY.
FT	DISULFID	813	833	BY SIMILIARITY.
FT	DISULFID	836	856	BY SIMILIARITY.
FT	DISULFID	874	894	BY SIMILIARITY.
FT	DISULFID	910	930	BY SIMILIARITY.
FT	DISULFID	946	1167	BY SIMILIARITY.
FT	CARBOHYD	151	151	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	316	316	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	455	455	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	582	582	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	708	708	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	936	936	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1067	1067	N-LINKED (GLCNAC. . .)
FT	CONFLICT	535	535	A -> V (IN REF. 3).
FT	CONFLICT	748	748	S -> T (IN REF. 3).
50	SEQUENCE	1170 AA:	129862 MW:	96F1BF53B89A051 CRC64;

Query Match	14.9%;	Score 143;	DB 1;	Length 1170;
Best Local Similarity	43.9%;	Pred. No. 0.0001;		
Matches	25;	Conservative	5;	Mismatches 27;
			Indels	0;
			Gaps	0;

Qy 117 GESHNNTKKKKDACIICECQDGVTCFEACPPATCAVPVNI PGACCPVCLORAE 173
| | | : | | | : | : | | | | | : | | | | :
Db 324 GRFAENETNVDSCIKCTCKKFKYVHQISCPATCADPWFVGECCSCSVHGEES 380

RESULT: 5	
TSP2_MOUSE	
ID	TSP2_MOUSE
STANDARD;	
PRT;	1172 AA

DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thrombospondin 2 precursor.
GN THBS2 OR TSP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147683; PubMed=1371115;
RA Lathery C.D., O'Rourke K., Wolf F.W., Katz R., Seidlin M.F.,
RA Dixit V.M.;
RT "Characterization of mouse thrombospondin 2 sequence and expression
during cell growth and development."
RL J. Biol. Chem. 267:3274-3281(1992).
RN [2]
RP SEQUENCE OF 1-873 FROM N.A.
RX MEDLINE=91302287; PubMed=1712771;
RA Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
RA Dixit V.M.;
RT "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
genome."
RL J. Biol. Chem. 266:12821-12824(1991).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: Homotrimer; disulfide-linked.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: Contains 1 VMRC domain.
CC -1- SIMILARITY: Contains 3 EGF-like domains.
CC -1- SIMILARITY: Contains 3 TSP type-1 domains.
CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

[illegible]

FT CONFLICT 383 383 N -> D (IN REF. 2).
 FT CONFLICT 573 573 Y -> H (IN REF. 2).
 FT CONFLICT 626 626 S -> C (IN REF. 2).
 SQ SEQUENCE 810 AA: 89606 MW: 549465EA37AEED0 CRC64:

Query Match 14.3%; Score 137.5; DB 1; Length 810;
 Best Local Similarity 30.8%; Pred. No. 0.0021;
 Matches 32; Conservative 16; Mismatches 43; Indels 13; Gaps 4;

OY 80 NDFREFV---LEMQTTIDLRTOIKLESRLSTTE-----CVDAGESHANNNTW-KKD 129
 DB 231 SDFSLVYGIMDLQELAKMTAKLNVAETRLQLENCHCEKTCQVSGLYRQDSWVDGND 290
 OY 130 ACTICECKGOVTCFEACPPATC---AVPVNIPGACCPVCLQK 170
 DB 291 HCRNCTCKSGAVECRMSPCLNCPSPDLPVHISGQCKVCRPK 334

RESULT 7
 NEIL_RAT
 ID NEIL_RAT STANDARD: PRT: 810 AA.
 AC 062919;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protein kinase C-binding protein NEIL1 precursor (NEIL-like protein 1).
 GN NEIL1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-Sprague-Dawley; TISSUE-Brain;
 RX MEDLINE=20017976; PubMed=10548494;
 RA Kuroda S., Oyatsu M., Kawakami M., Kanayama N., Tanizawa K., Saito N.,
 RA Abe T., Matsubashi S., Ting K.;
 RT "Biochemical characterization and expression analysis of neural
 RT thrombospondin-1-like proteins NEIL1 and NEIL2.";
 RL Biochem. Biophys. Res. Commun. 265:79-86(1999).

CC -1- SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -1- SIMILARITY: Contains 5 WFC domains.
 CC -1- SIMILARITY: Contains 6 EGF-like domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.cn/announce/>
 CC or send an email to license@sdb-sdb.ch).

DR EMBL: U48246; AAC72252.1; -;
 DR PIR: T10756; T10756.
 DR HSSP: P35555; IEMN.
 DR InterPro: IPR00152; Asx_hydroxyl.
 DR InterPro: IPR00181; EGF_CA.
 DR InterPro: IPR006209; EGF-like.
 DR InterPro: IPR001791; Laminin_G.
 DR InterPro: IPR003129; TSPN.
 DR InterPro: IPR001007; WFC_C.
 DR Pfam: PF00008; EGF_5.
 DR Pfam: PF02210; TSPN; 1.
 DR Pfam: PF00093; WFC; 2.
 DR SMART: SM00179; EGF_CA; 2.
 DR SMART: SM00283; LamG; 1.
 DR SMART: SM00210; TSPN; 1.
 DR SMART: SM00214; WFC; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 3.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 3.

DR PROSITE: PS01187; EGF_CA; 3.
 DR PROSITE: PS01208; WFC_1; 2.
 DR PROSITE: PS00184; WFC_2; 2.
 KW Glycoprotein; EGF-like domain; Repeat; Signal.
 FT SIGNAL 1
 FT CHAIN 17 810
 FT DOMAIN 81 230
 FT DOMAIN 271 332
 FT DOMAIN 335 390
 FT DOMAIN 391 433
 FT DOMAIN 434 475
 FT DOMAIN 476 516
 FT DOMAIN 515 547
 FT DOMAIN 549 595
 FT DOMAIN 596 631
 FT DOMAIN 632 687
 FT DOMAIN 692 750
 FT DOMAIN 752 807
 FT DISULFID 395 407
 FT DISULFID 401 416
 FT DISULFID 418 432
 FT DISULFID 438 451
 FT DISULFID 445 460
 FT DISULFID 462 474
 FT DISULFID 480 493
 FT DISULFID 487 502
 FT DISULFID 504 515
 FT DISULFID 519 529
 FT DISULFID 523 535
 FT DISULFID 537 546
 FT DISULFID 553 566
 FT DISULFID 560 575
 FT DISULFID 577 594
 FT DISULFID 600 613
 FT DISULFID 607 622
 FT DISULFID 624 630
 FT CARBOHYD 40 40
 FT CARBOHYD 53 53
 FT CARBOHYD 83 83
 FT CARBOHYD 224 224
 FT CARBOHYD 294 294
 FT CARBOHYD 372 372
 FT CARBOHYD 511 511
 FT CARBOHYD 562 562
 FT CARBOHYD 609 609
 FT CARBOHYD 708 708
 SQ SEQUENCE 810 AA: 89212 MW: 46F09C466AF9AB0B CRC64:

Query Match 14.0%; Score 134.5; DB 1; Length 810;
 Best Local Similarity 32.0%; Pred. No. 0.0038;
 Matches 31; Conservative 14; Mismatches 47; Indels 5; Gaps 3;

OY 78 GINDFREFVLEMQTTIDLRTOIKLESRLSTTECVDAAGESHANNNTW-KKDACTICEC 136
 DB 239 GIMDLQELAKMTAKLNVAETRLQLENCHCEKTC-QVSGLYRQDSWVDGNDGNCNCIC 297
 OY 137 KDGQVTCFEACPPATC---AVPVNIPGACCPVCLQK 170
 DB 298 KSGAVECRMSPCLNCPSPDLPVHISGQCKVCRPK 334

RESULT 8
 NRLL_HUMAN
 ID NRLL_HUMAN STANDARD: PRT: 450 AA.
 AC Q9BU04; Q9Y3H7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Neuralin precursor (Ventropilin).
 GN NRNLN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=22388257, PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Usdin T.B., Tashyuk S., Carrinci P., Prange C.,
 Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN
 RP SEQUENCE OF 197-450 FROM N.A.
 RA Pearce A.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Seems to antagonize the function of BMP-4 by binding to
 it and preventing its interaction with receptors (by similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -1- SIMILARITY: Contains 3 WFC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: BC002909; AA02909.1; -;
 DR EMBL: AL049176; CAB42789.1; -;
 DR InterPro: IPR001007; WFC_C.
 DR Pfam: PF00093; WFC; 3.
 DR SMART: SM00214; WFC; 3.
 DR PROSITE: PS01208; WFC_1; 3.
 DR PROSITE: PS0184; WFC_2; 3.
 KW Developmental protein; Repeat; Glycoprotein; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 450
 FT DOMAIN 29 94
 FT DOMAIN 107 173
 FT DOMAIN 252 317
 FT SITE 173 175
 FT CARBOHYD 112 112
 FT CARBOHYD 285 285
 FT CONFLICT 323 323
 FT CONFLICT 323 323
 SQ SEQUENCE 450 AA: 51167 MW: D5F39AB9E2EE3E71 CRC64;
 Query Match 13.4%; Score 129; DB 1; Length 450;
 Best Local Similarity 51.4%; Pred. No. 0.0006;
 Matches 19; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

AC Q902D5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neuralin precursor (Ventrolin).
 GN NR1M1 OR VOPF.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN (1)
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND FUNCTION.
 RX MEDLINE=21334726; PubMed=11441185;
 RA Sakuta H., Suzuki R., Takahashi H., Kato A., Shintani T.,
 Iemura S.-I., Yamamoto T.S., Ueno N., Noda M.;
 RT "Ventrolin: a BMP-4 antagonist expressed in a double-gradient
 pattern in the retina";
 RL Science 293:111-115(2001).
 CC -1- FUNCTION: Seems to antagonize the function of BMP-4 by binding to
 it and preventing its interaction with receptors.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Mainly expressed in the ventral retina.
 CC -1- SIMILARITY: Contains 3 WFC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF257352; AA073359.1; -;
 DR InterPro: IPR001007; WFC_C.
 DR Pfam: PF00093; WFC; 3.
 DR SMART: SM00214; WFC; 3.
 DR PROSITE: PS01208; WFC_1; 3.
 DR PROSITE: PS0184; WFC_2; 3.
 KW Developmental protein; Repeat; Glycoprotein; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 456
 FT SITE 181 183
 FT CARBOHYD 120 120
 FT CARBOHYD 295 295
 SQ SEQUENCE 456 AA: 52399 MW: D54F6685F3839BC6 CRC64;
 Query Match 13.4%; Score 129; DB 1; Length 456;
 Best Local Similarity 48.6%; Pred. No. 0.00061;
 Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

OY 131 CTICCKDGOVTCFEACPPATCAVPVNI PGACCPVC 167
 DB 144 CAQCSGSEGNVYCGKTCPKLTCFVPSVSCCVC 180
 RESULT 10
 ECM2_HUMAN
 ID ECM2_HUMAN STANDARD; PRT; 699 AA.
 AC 094769;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Extracellular matrix protein 2 precursor (Matrix glycoprotein
 SCL/ECM2).
 GN ECM2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99009324; PubMed=9790758;

RA Nishiu J., Tanaka T., Nakamura Y.;
 RT Identification of a novel gene (ECM2) encoding a putative
 RT extracellular matrix protein expressed predominantly in adipose and
 RT female-specific tissues and its chromosomal localization to 9q22.3.";
 RL Genomics 52:378-381(1998).
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -1- TISSUE SPECIFICITY: Expressed predominantly in adipose tissue as
 CC well as female-specific organs such as mammary gland, ovary, and
 CC uterus.
 CC -1- SIMILARITY: BELONGS TO THE SMALL LEUCINE-RICH PROTEOGLYCAN
 CC (SLRPs) FAMILY.
 CC -1- SIMILARITY: Contains 1 WFEC domain.
 CC -1- SIMILARITY: Contains 12 leucine-rich (LRR) repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB011792; BAA3958.1; -.
 CC Genew: HGNC:3154; ECM2.
 DR MIM: 603479; -.
 DR GO: GO:0005578; C:extracellular matrix; TAS.
 DR GO: GO:0005178; F:integrin binding activity; TAS.
 DR GO: GO:0007160; P:cell matrix adhesion; TAS.
 DR Interpro: IPR001611; LRR.
 DR Interpro: IPR003591; LRR_Typ.
 DR Interpro: IPR001007; WFEC_C.
 DR Pfam: PF00560; LRR; 10.
 DR Pfam: PF00093; VWC; 1.
 DR PRINTS: PRO0019; LEURICHRPT.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; WFEC_1; 1.
 DR PROSITE: PS0184; WFEC_2; 1.
 KW Leucine-rich repeat; Repeat; Glycoprotein; Extracellular matrix;
 KW Signal.
 FT CHAIN 1 20 POTENTIAL.
 FT SIGNAL 1 20 EXTRACELLULAR MATRIX PROTEIN 2.
 FT DOMAIN 101 158 WFEC.
 FT REPEAT 335 355 LRR-S 1.
 FT REPEAT 356 379 LRR-T 1.
 FT REPEAT 382 405 LRR-T 2.
 FT REPEAT 406 426 LRR-S 2.
 FT REPEAT 427 450 LRR-T 3.
 FT REPEAT 451 476 LRR-T 4.
 FT REPEAT 477 497 LRR-S 3.
 FT REPEAT 524 547 LRR-T 5.
 FT REPEAT 547 566 LRR-T 6.
 FT REPEAT 621 643 LRR-S 4.
 FT REPEAT 644 672 LRR-T 7.
 FT REPEAT 673 699 LRR-T 8.
 FT DOMAIN 270 281 POLY-GLU.
 FT SITE 294 296 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 378 378 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 506 506 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 699 AA: 79789 MW: E44E76A40A5C2742 CRC64;
 Query Match 13.2%; Score 127; DB 1; Length 699;
 Best Local Similarity 25.6%; Pred. No. 0.0014;
 Matches 40; Conservative 14; Mismatches 62; Indels 40; Gaps 6;

OY 1332 TICECKDGVTCFEVACPPATCAVPIVNGACPC 167
 DB 122 TITCLSDGRVLCDEFCHQRCPQYIVPGECCPC 157
 RESULT 11
 ID NR1L MOUSE STANDARD; PRT; 447 AA.
 AC Q920C1; Q924K0; Q9EP29;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neuralin precursor (Ventropilin).
 GN NR1L.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
 RC STRAIN=C57BL/6J;
 RX MEDLINE=21334726; PubMed=11441185;
 RA Sakuta H., Suzuki R., Takahashi H., Kato A., Shintani T.,
 RA Iemura S.-I., Yamamoto T.S., Ueno N., Noda M.;
 RT "Ventropilin: a BMP-4 antagonist expressed in a double-gradient
 RT pattern in the retina.";
 RL Science 293:111-115(2001).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM BETA).
 RX MEDLINE=20568969; PubMed=1118896;
 RA Coffinier C.C., Tran U., Larrain J., De Robertis E.M.;
 RT "Neuralin-1 is a novel chordin-related molecule expressed in the mouse
 RT neural plate.";
 RL Mech. Dev. 100:119-122(2001).
 CC -1- FUNCTION: Seems to antagonize the function of BMP-4 by binding to
 CC it and preventing its interaction with receptors (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms-2:
 CC Name=Alpha:
 CC IsoId=Q920C1-1; Sequence=Displayed;
 CC Name=Beta:
 CC IsoId=Q920C1-2; Sequence=VSP_001076, VSP_001077;
 CC -1- TISSUE SPECIFICITY: Expressed in heart, brain, lung, liver, kidney
 CC and testis.
 CC -1- DEVELOPMENTAL STAGE: Expression starts in the neural plate at mid-
 CC gastrulation. Later on its expression becomes restricted to
 CC discrete regions of the central nervous system and to derivatives
 CC of the neural crest cells.
 CC -1- SIMILARITY: Contains 3 WFEC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF21853; AAK95586.1; -.
 CC EMBL: AF296451; AAK71523.1; -.
 CC EMBL: AF305714; AAG27460.1; -.
 CC MGD: MGI:1933172; Nr1l.
 DR Interpro: IPR001007; WFEC_C.
 DR Pfam: PF00093; VWC; 3.
 DR SMART: SM00214; VWC; 3.
 DR PROSITE: PS01208; WFEC_1; 3.
 DR PROSITE: PS0184; WFEC_2; 3.
 KW developmental protein; Repeat; Glycoprotein; signal;
 KW Alternative splicing.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 447 NEURALIN.
 FT DOMAIN 30 95 WFEC 1.

```

FT DOMAIN 108 174 VMEC 2.
FT DOMAIN 253 318 VMEC 3.
FT SITE 174 176 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 319 333 EEPSPNPFDSKSPFC -> GKAKAGLAGPAFG (in
FT ISOFORM Beta).
FT VARSPPLIC 334 447 /FTID-VSP_001076.
FT VARSPPLIC 447 447 Missing (in isoform Beta).
FT CONFLICT 211 211 /FTID-VSP_001077.
FT CONFLICT 211 211 N -> S (IN REF. 2).
SQ SEQUENCE 447 AA; 50732 MW; 94C99025A97545D CRC64;
Query Match 13.0%; Score 125; DB 1; Length 447;
Best Local Similarity 48.6%; Pred. No. 0.0013;
Matches 18; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 131 CTICECKDGVYCFEACPPATCAVPYNTFGACCPYC 167
DB 137 CSQCSGSEGVYCGKTCPLKCAFPVSVSDSCRCVC 173

RESULT 12
TSPL_HUMAN STANDARD; PRT; 1170 AA.
ID TSPL_HUMAN
AC P07996; Q15667;
DC 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Thrombospondin 1 precursor.
GN THBS1 OR TSPL OR TSP.
OS Homo sapiens (Human).
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=87057617; PubMed=2430973;
RA Lawler J., Hynes R.O.;
RT "The structure of human thrombospondin, an adhesive glycoprotein with
RT multiple calcium-binding sites and homologies with several different
RT proteins."
RL J. Cell Biol. 103:1635-1648(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89139590; PubMed=2918029;
RA Hennessy S.W., Frazier B.A., Kim D.D., Deckwerth T.L.,
RA Baumgartel D.M., Rotwein P., Frazier W.A.;
RT "Complete thrombospondin mRNA sequence includes potential regulatory
RT sites in the 3' untranslated region."
RL J. Cell Biol. 108:729-736(1989).
RN [3]
RP SEQUENCE OF 1-397 FROM N.A.
RX MEDLINE=87157592; PubMed=3030396;
RA Kobayashi S., Eden-McCutchan F., Framson P., Bornstein P.;
RT "Partial amino acid sequence of human thrombospondin as determined by
RT analysis of cDNA clones: homology to malarial circumsporozoite
RT proteins."
RL Biochemistry 25:8418-8425(1986).
RN [4]
RP SEQUENCE OF 1-374 FROM N.A.
RX MEDLINE=86287276; PubMed=3461443;
RA Dixit V.M., Hennessy S.W., Grant G.A., Rotwein P., Frazier W.A.;
RT "Characterization of a cDNA encoding the heparin and collagen binding
RT domains of human thrombospondin."
RL Proc. Natl. Acad. Sci. U.S.A. 83:5449-5453(1986).
RN [5]
RP SEQUENCE OF 1-166 FROM N.A.
RX MEDLINE=89291870; PubMed=2544587;
RA Laherty C.D., Gierman T.M., Dixit V.M.;
RT "Characterization of the promoter region of the human thrombospondin
RT gene. DNA sequences within the first intron increase transcription."

```

```

RL J. Biol. Chem. 264:11222-11227(1989).
RN [6]
RP SEQUENCE OF 1028-1170 FROM N.A.
RA la Fleur M., Jobin C., Gauthier J., Kreis C.G.;
RN Submitted (xxx1992) to the EMBL/Genbank/DBJ databases.
RN [7]
RP CARBOHYDRATE-LINKAGE SITES TRP-385; SER-394; TRP-438; TRP-441;
RP THR-450; TRP-498 AND THR-507.
RC TISSUE=platelet;
RX MEDLINE=21125860; PubMed=11067851;
RA Hofsteenge J., Huwiler K.G., Macek B., Hess D., Lawler J.,
RA Mosher D.F., Peter-Katalinic J.;
RT "C-mannosylation and O-fucosylation of the thrombospondin type 1
RT module."
RL J. Biol. Chem. 276:6485-6498(2001).
RN [8]
RP THROMBOSPONDIN DOMAIN DISULFIDE BRIDGES.
RX MEDLINE=22338361; PubMed=12450399;
RA Huwiler K.G., Vestling M.M., Annis D.S., Mosher D.F.;
RT "Biophysical characterization, including disulfide bond assignments,
RT of the anti-angiogenic type 1 domains of human thrombospondin-1."
RL Biochemistry 41:14329-14339(2002).
CC -1- FUNCTION: Adhesive glycoprotein that mediates cell-to-cell and
CC cell-to-matrix interactions. Can bind to fibrinogen, fibronectin,
CC laminin, type V collagen and integrins alpha-V/beta-1, alpha-
CC V/beta-3 and alpha-ITB/beta-3.
CC -1- SUBUNIT: Homotrimer; disulfide-linked.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: Contains 1 VMEC domain.
CC -1- SIMILARITY: Contains 3 EGF-like domains.
CC -1- SIMILARITY: Contains 7 TSP type-1 domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M25631; AAA6741.1; -
DR EMBL: X04665; CAA28370.1; -
DR EMBL: X14787; CAA32889.1; -
DR EMBL: M14326; CAA61237.1; ALT_SEQ.
DR EMBL: J04835; AAA61178.1; -
DR EMBL: M99425; AAB59366.1; -
DR PIR: A26155; TSHNP1.
DR PDB: 1LSL; 18-DEC-02.
DR GlycoSuiteDB: P07996; -
DR GeneW: HGNC:11785; THBS1.
DR MIM: 188060; -
DR GO: GO:0004866; F:endorpeptidase inhibitor activity; TAS.
DR GO: GO:0004871; F:signal transducer activity; TAS.
DR GO: GO:0007275; P:development; TAS.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR003367; TSP-3.
DR InterPro: IPR003129; TSPN.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF00008; EGF_2.
DR Pfam: PF00090; TSP_1; 3.
DR Pfam: PF00093; VWC_1.
DR Pfam: PF02210; TSPN; 1.
DR Pfam: PF02412; TSP_3; 8.
DR SMART: SM00181; EGF; 3.
DR SMART: SM00209; TSP1; 3.
DR SMART: SM00210; TSPN; 1.
DR SMART: SM00214; VWC; 1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.

```


DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; WFRC_1; 1.
 DR PROSITE; PS0184; WFRC_2; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 EGF-like domain; Signal; 3D-structure.
 FT CHAIN 1 18
 FT SIGNAL 19 1170
 FT DOMAIN 19 232
 FT DOMAIN 24 221
 FT DOMAIN 316 373
 FT DOMAIN 379 429
 FT DOMAIN 435 490
 FT DOMAIN 492 547
 FT DOMAIN 549 587
 FT DOMAIN 588 645
 FT DOMAIN 646 690
 FT DOMAIN 723 758
 FT DOMAIN 759 781
 FT DOMAIN 782 817
 FT DOMAIN 818 840
 FT DOMAIN 841 878
 FT DOMAIN 879 914
 FT DOMAIN 915 950
 FT DOMAIN 951 1170
 FT SITE 928 928
 FT DISULFID 270 270
 FT DISULFID 274 274
 FT DISULFID 391 423
 FT DISULFID 395 428
 FT DISULFID 406 413
 FT DISULFID 447 484
 FT DISULFID 451 489
 FT DISULFID 462 474
 FT DISULFID 504 541
 FT DISULFID 508 546
 FT DISULFID 519 531
 FT DISULFID 551 562
 FT DISULFID 556 572
 FT DISULFID 575 586
 FT DISULFID 592 608
 FT DISULFID 599 617
 FT DISULFID 620 644
 FT DISULFID 650 663
 FT DISULFID 657 676
 FT DISULFID 678 689
 FT DISULFID 705 713
 FT DISULFID 718 738
 FT DISULFID 754 774
 FT DISULFID 777 797
 FT DISULFID 813 833
 FT DISULFID 836 856
 FT DISULFID 874 894
 FT DISULFID 910 930
 FT DISULFID 946 1167
 FT CARBOHYD 248 248
 FT CARBOHYD 360 360
 FT CARBOHYD 385 385
 FT CARBOHYD 394 394
 FT CARBOHYD 438 438
 FT CARBOHYD 441 441
 FT CARBOHYD 450 450
 FT CARBOHYD 498 498
 FT CARBOHYD 507 507
 FT CARBOHYD 708 708
 FT CARBOHYD 1067 1067
 THROMBOSPONDIN 1.
 HEPARIN-BINDING (POTENTIAL).
 TSP N-TERMINAL.
 WFRC.
 TSP TYPE-1 1.
 TSP TYPE-1 2.
 TSP TYPE-1 3.
 EGF-LIKE 1.
 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 EGF-LIKE 3.
 TSP TYPE-3 1.
 TSP TYPE-3 2.
 TSP TYPE-3 3.
 TSP TYPE-3 4.
 TSP TYPE-3 5.
 TSP TYPE-3 6.
 TSP TYPE-3 7.
 C-TERMINAL.
 CELL ATTACHMENT SITE (POTENTIAL).
 INTERCHAIN (PROBABLE).
 INTERCHAIN (PROBABLE).

FT CONFLICT 84 84 T -> A (IN REF. 2, 3 AND 4).
 FT CONFLICT 523 523 T -> A (IN REF. 2).
 SQ SEQUENCE 1170 AA, 129412 MW, 69B3EDE5AE3A395E CXC6;
 Query Match 13.0%; Score 125; DB 1; Length 1170;
 Best Local Similarity 29.4%; Pred. No. 0.0037;
 Matches 25; Conservative 13; Mismatches 47; Indels 0; Gaps 0;
 Db 348 CRRVSCIPMCNSNATVPDGCPCRC 372
 143 CFVEACPPATCAVPVNPACPCPC 167
 288 RTVTLTLDSDIRRVTEENKELANLRPPCLYHNGVQRNNEWTDSCTECHCONSVTI 347
 83 REYVLEKQKITTLRLQIKLESRLSTTECVDAAGESHANNTWKADCTICCKGQVT 142
 RESULT 13
 TSP1_MOUSE STANDARD; PRT; 1170 AA.
 ID TSP1_MOUSE
 AC P35441;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Thrombospondin 1 precursor.
 GN THBS1 OR TSP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lawler J., Duquette M., Ferro P., Copeland N.G., Gilbert D.J.,
 RA Denkins N.A.;
 RT "Characterization of the murine thrombospondin gene.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92147683; PubMed=1371115;
 RA Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
 RA Dixit V.M.;
 RT "Characterization of mouse thrombospondin 2 sequence and expression
 during cell growth and development.";
 RL J. Biol. Chem. 267:3274-3281(1992).
 RN [3]
 RP SEQUENCE OF 1-490 FROM N.A.
 RX MEDLINE=90375546; PubMed=2398070;
 RX Bornstein P., Alfi D., Devarayalu S., Framson P., Li P.;
 RT "Characterization of the mouse thrombospondin gene and evaluation of
 the role of the first intron in human gene expression.";
 RL J. Biol. Chem. 265:16691-16698(1990).
 CC -1- FUNCTION: Adhesive glycoprotein that mediates cell-to-cell and
 cell-to-matrix interactions. Can bind to fibrinogen, fibronectin,
 laminin, type V collagen and integrins alpha-V/beta-1, alpha-
 V/beta-3 and alpha-IIb/beta-3.
 CC -1- SUBUNIT: Homotrimer; disulfide-linked.
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -1- SIMILARITY: Contains 1 WFRC domain.
 CC -1- SIMILARITY: Contains 3 EGF-like domains.
 CC -1- SIMILARITY: Contains 3 TSP type-1 domains.
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M62470; AAA50611.1; -

DR EMBL; M62450; AAA50611.1; JOINED.
 DR EMBL; M62451; AAA50611.1; JOINED.
 DR EMBL; M62452; AAA50611.1; JOINED.
 DR EMBL; M62453; AAA50611.1; JOINED.
 DR EMBL; M62454; AAA50611.1; JOINED.
 DR EMBL; M62455; AAA50611.1; JOINED.
 DR EMBL; M62456; AAA50611.1; JOINED.
 DR EMBL; M62457; AAA50611.1; JOINED.
 DR EMBL; M62458; AAA50611.1; JOINED.
 DR EMBL; M62459; AAA50611.1; JOINED.
 DR EMBL; M62460; AAA50611.1; JOINED.
 DR EMBL; M62461; AAA50611.1; JOINED.
 DR EMBL; M62462; AAA50611.1; JOINED.
 DR EMBL; M62463; AAA50611.1; JOINED.
 DR EMBL; M62464; AAA50611.1; JOINED.
 DR EMBL; M62465; AAA50611.1; JOINED.
 DR EMBL; M62466; AAA50611.1; JOINED.
 DR EMBL; M62467; AAA50611.1; JOINED.
 DR EMBL; M62468; AAA50611.1; JOINED.
 DR EMBL; M62469; AAA50611.1; JOINED.
 DR EMBL; M87276; AAA53063.1; -
 DR EMBL; J05606; AAA40431.1; -
 DR EMBL; J05605; AAA40431.1; JOINED.
 DR PIR; A40558; A40558.
 DR MGD; MGI:98737; Thbs1.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR003367; TSP3.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR001007; WVC_Like.
 DR Pfam; PF00090; tsf_1; 3.
 DR Pfam; PF02412; tsf_3; 8.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00093; WVC; 1.
 DR SMART; SM00181; EGF; 3.
 DR SMART; SM00209; TSP1; 3.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00214; WVC; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; WVC_Like; 1.
 DR PROSITE; PS50184; WVC_2; 1.
 DR Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 EGF-like domain; signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1170
 FT DOMAIN 19 232
 FT DOMAIN 24 221
 FT DOMAIN 316 373
 FT DOMAIN 379 429
 FT DOMAIN 435 490
 FT DOMAIN 492 547
 FT DOMAIN 549 587
 FT DOMAIN 588 645
 FT DOMAIN 646 690
 FT DOMAIN 723 758
 FT DOMAIN 759 781
 FT DOMAIN 782 817
 FT DOMAIN 818 840
 FT DOMAIN 841 878
 FT DOMAIN 879 914
 FT DOMAIN 915 950
 FT DOMAIN 951 1170
 FT SITE 926 928
 FT DISULFID 270 274
 FT DISULFID 274 274
 FT DISULFID 391 423
 FT DISULFID 395 428
 FT DISULFID 406 413

FT DISULFID 447 484 BY SIMILARITY.
 FT DISULFID 451 489 BY SIMILARITY.
 FT DISULFID 462 479 BY SIMILARITY.
 FT DISULFID 504 541 BY SIMILARITY.
 FT DISULFID 508 546 BY SIMILARITY.
 FT DISULFID 519 531 BY SIMILARITY.
 FT DISULFID 551 562 BY SIMILARITY.
 FT DISULFID 556 572 BY SIMILARITY.
 FT DISULFID 575 586 BY SIMILARITY.
 FT DISULFID 592 608 BY SIMILARITY.
 FT DISULFID 599 617 BY SIMILARITY.
 FT DISULFID 620 644 BY SIMILARITY.
 FT DISULFID 650 663 BY SIMILARITY.
 FT DISULFID 657 676 BY SIMILARITY.
 FT DISULFID 678 689 BY SIMILARITY.
 FT DISULFID 705 713 BY SIMILARITY.
 FT DISULFID 718 728 BY SIMILARITY.
 FT DISULFID 754 774 BY SIMILARITY.
 FT DISULFID 777 797 BY SIMILARITY.
 FT DISULFID 813 833 BY SIMILARITY.
 FT DISULFID 836 856 BY SIMILARITY.
 FT DISULFID 874 894 BY SIMILARITY.
 FT DISULFID 910 930 BY SIMILARITY.
 FT DISULFID 946 1167 BY SIMILARITY.
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1067 1067 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1025 1025 F -> L (IN REF. 2).
 SQ SEQUENCE 1170 AA; 129646 MW; 0435493615E7F06 CRC64;
 Query Match 13.0%; Score 125; DB 1; Length 1170;
 Best Local Similarity 29.4%; Pred. No. 0.0037;
 Matches 25; Conservative 13; Mismatches 47; Indels 0; Gaps 0;
 QY 83 REFVLEMQKTTDLTQIKLESRLSTECVAGSESHANNKMKKADCTICECDGQVT 142
 DB 288 RTIVTTLDSDIRKVTENNELVSELKRPPLCFHNGVQYKNNEWTVDSCTECHCONSVTI 347
 QY 143 CFVEACPATCAVPNIPGACCPVC 167
 DB 348 CKKVCSPIMPCSNATVPDGECCPRC 372
 RESULT 14
 ID CHRD. HUMAN STANDARD; PRT: 955 AA.
 AC Q9H2X0; Q95254; Q9H2D3; Q9H2W8; Q9H2W9; Q9P0Z2; Q9P0Z3; Q9P0Z4;
 AC Q9P0Z5;
 DT 16-OCT-2001 (rel. 40; Created)
 DT 16-OCT-2001 (rel. 40; Last sequence update)
 DT 28-FEB-2003 (rel. 41; Last annotation update)
 DE Chordin precursor.
 GN CHRD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).
 RX MEDLINE=21366001; PubMed=11472837;
 RA Millet C., Lemaire P., Orsetti B., Guglielmi P., Francois V.;
 RT "The human chordin gene encodes several differentially expressed
 RT spliced variants with distinct BMP opposing activities.";
 Mech. Dev. 106:85-96(2001).
 RP [2]
 RP SEQUENCE OF 115-955 FROM N.A. (ISOFORM 5).
 RX MEDLINE=99000848; PubMed=9782094;
 RA Pappano W.N., Scott I.C., Clark T.G., Eddy R.L., Shows T.B.,
 RA Greenspan D.S.;
 RT "Coding sequence and expression patterns of mouse chordin and mapping
 RT of the cognate mouse chrd and human CHRD genes.";
 Genomics 52:236-239(1998).
 RL


```

RP SEQUENCE OF 288-297 AND 606-617.
RC TISSUE-BONE.
RA MEDLINE-94237164; PubMed-8181482.
RA Botradl-Anelli M., Rousseau J.C., Klemm J.P., Champliand M.F.,
RA Mourad M.M., Bernillon J., Wallach J.M., van der Rest M.;
RA "Diversity in the processing events at the N-terminus of type-V
RA collagen";
RA Eur. J. Biochem. 221:987-995(1994).
RN [7]
RP DISEASE.
RX MEDLINE-96087576; PubMed-9425231;
RA Michalickova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
RT "Mutations of the alpha2(V) chain of type V collagen impair matrix
RT assembly and produce Ehlers-Danlos syndrome type I.";
RL Hum. Mol. Genet. 7:249-255(1998).
RN [8]
RP VARIANT EDS-III ARG-960.
RX MEDLINE-98455031; PubMed-9783710;
RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
RA Burrows N.P.;
RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
RT II.";
RL J. Med. Genet. 35:846-848(1998).
CC -I- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGENS). IT IS A MINOR CONNECTIVE TISSUE
CC COMPONENT OF NEARLY UBQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
CC TO DNM, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
CC -I- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN
CC MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
CC ONE ALPHA 3(V) CHAINS IN PLACENTA.
CC -I- PMW: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -I- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
CC gravels. EDS-I is an autosomal dominant connective-tissue disorder
CC characterized by loose-jointedness and fragile, velvety,
CC stretchable, bruisable skin that heals with peculiar 'cigarette-
CC paper' scars. Inheritance is autosomal dominant.
CC -I- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
CC syndrome mitis. Inheritance is autosomal dominant.
CC -I- SIMILARITY: Contains 1 WWC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04478; AAA51859.1; -
DR EMBL; X04758; CA28454.1; -
DR EMBL; M11718; AAA52058.1; -
DR PIR; A31427; CGH2V.
DR PDB; 1A9A; 18-NOV-98.
DR Genew; HGNC:2210; COL5A2.
DR MIM; 120190; -
DR MIM; 130000; -
DR MIM; 130010; -
DR GO; GO:0005588; C:collagen type V; TAS.
DR GO; GO:0005202; F:collagen; TAS.
DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; FibCollagen_C.
DR InterPro; IPR001007; WWC_C.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; WWC; 1.
DR ProDom; PD000007; Ctg_helix; 5.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLF1; 1.
DR SMART; SM00214; WWC; 1.

```

[illegible]

Search completed: July 24, 2003, 12:52:23
Job time : 6.86524 secs

```

DR PROSITE: PS01208; WMEC_1; 1.
DR PROSITE: PS50184; WMEC_2; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Ehlers-Danlos syndrome;
KW Disease mutation; 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 1226 COLLAGEN ALPHA 2(V) CHAIN.
FT PROPEP 1227 1496 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 39 97 WMEC.
FT MOD_RES 290 290 HYDROXYLATION.
FT MOD_RES 293 293 HYDROXYLATION.
FT MOD_RES 296 296 HYDROXYLATION.
FT MOD_RES 608 608 HYDROXYLATION.
FT MOD_RES 614 614 HYDROXYLATION.
FT VARIANT 960 960 G -> R (in EDS-II).
FT CONFLICT 292 292 /FTID=VAR_013568.
FT CONFLICT 1418 1418 K -> P (in REF. 6).
FT CONFLICT 1438 1438 K -> T (in REF. 3).
FT CONFLICT 1460 1460 F -> S (in REF. 3).
FT CONFLICT 1496 1496 E -> Q (in REF. 4).
FT CONFLICT 1496 1496 V -> A (in REF. 4).
SO SEQUENCE 1496 AA; 144720 MM; 82827C17A8644F5A CRC64;

Query Match
Best Local Similarity 12.8%; Score 123.5; DB 1; Length 1496;
Matches 23; Conservative 5; Mismatches 25; Indels 1; Gaps 1.

OY 117 GESHANNTRKWKDADCTICECKDQGVTC-FVEACPPATCAVPVNIIGACCPVCLQ 169
DB 45 GQWTLNRDIWKRPAPCQICVCDNGAILCDKIECQDVLDCADDPVTPRGECPCVCSQ 98

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:46:04 ; Search time 8.6358 Seconds
(without alignments)
1959.945 Million cell updates/sec

Title: US-09-884-319a-2
Perfect score: 962
Sequence: 1 IPRVDLRVWQDCCECDRTTRG.....VNIPGACCPVCLQKRAEKP 176

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PTR_76:*
1: ptr1:*
2: ptr2:*
3: ptr3:*
4: ptr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	153.5	16.0	1535	2 S46224	peroxidase - fruit
2	148	15.4	1178	1 A39804	thrombospondin pre
3	147	15.3	941	1 A51195	chordin precursor
4	147	15.3	1172	1 TSHUP2	thrombospondin 2 p
5	139.5	14.5	3198	2 A43426	collagen alpha 2 f
6	139	14.4	1172	2 A42587	thrombospondin 2 p
7	137	14.2	1376	2 S63986	collagen alpha 5 c
8	134.5	14.0	810	2 T10756	Nel-homolog protei
9	134.5	14.0	1822	2 S63985	collagen alpha 2 c
10	132.5	13.8	1497	2 T49607	procollagen type V
11	125	13.0	1170	1 TSHUP1	thrombospondin 1 p
12	125	13.0	1170	2 A40558	thrombospondin 1 p
13	123.5	12.8	1496	1 CGHU2V	collagen alpha 2V
14	113	11.7	206	2 S18250	collagen alpha 1(I
15	111	11.5	1492	2 A40333	collagen alpha 1(I
16	109	11.3	1486	1 B40333	collagen alpha 1(I
17	108	11.2	835	2 JP0076	collagen alpha 1(I
18	108	11.2	1466	1 CGHU7L	collagen alpha 1(I
19	107	11.1	153	2 A27179	collagen alpha 1(I
20	107	11.1	1464	1 CGHU15	collagen alpha 1(I
21	106	11.0	1038	2 T13177	collagen alpha 1(I
22	105.5	11.0	1487	2 B41182	collagen alpha 1(I
23	105	10.9	884	2 T18649	collagen alpha 1(I
24	104	10.8	615	2 A05269	collagen alpha 1(I
25	104	10.8	886	2 T50694	collagen alpha 1(I
26	103.5	10.7	194	2 T14746	hypothetical prote
27	102.5	10.7	3133	2 S52093	hemocytin - silkwo
28	102	10.6	98	2 I49562	collagen alpha 1(I
29	102	10.6	1464	2 S59856	collagen alpha 1(I

30	99	10.3	1487	1 CGHU6C	collagen alpha 1(I
31	97	10.1	114	2 A54663	seminal plasma pro
32	94	9.8	488	2 A27353	collagen alpha 1(I
33	92	9.6	114	2 A34567	beta-microseminopr
34	92	9.6	739	2 E86434	protein F17F8.27 l
35	91	9.5	712	2 S68306	pol polyprotein, t
36	88	9.1	91	2 S41663	beta-microseminopr
37	87.5	9.1	1034	2 JC5598	mucin - rat
38	87	9.0	5376	2 T42215	zonadhesin - mouse
39	86.5	9.0	1700	2 S08167	Baldiani ring 3 pr
40	84.5	8.8	249	2 T24604	hypothetical prote
41	84.5	8.8	422	2 D86446	hypothetical prote
42	84.5	8.8	940	2 H86420	probable receptor
43	84.5	8.8	1453	2 S21626	collagen alpha 1(I
44	83.5	8.7	2813	1 VWHU	von Willebrand fac
45	82.5	8.6	376	2 JC4892	L-selectin precurs

ALIGNMENTS

RESULT 1

S46224 peroxidase - fruit fly (Drosophila sp.)

C:Species: Drosophila sp.

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Oct-2000

C:Accession: S46224

R:Nelson, R.E.; Fessler, L.I.; Takagi, Y.; Blumberg, B.; Keene, D.R.; Olson, P.F.; Pa

EMBO J. 13, 3438-3447, 1994

A:Title: Peroxidase: a novel enzyme-matrix protein of Drosophila development.

A:Reference number: S46224; MID:94341255; PMID:8062820

A:Accession: S46224

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1535 <NEU>

A:Cross-references: GB:011052; NID:9531384; PIDN:AAA61568.1; PID:9531385

A:Superfamily: peroxidase; myeloperoxidase homology; proteoglycan amino-terminal hom

F:19-44/Domain: proteoglycan amino-terminal homology <PAH4>

F:661-1350/Domain: myeloperoxidase homology <MPX>

Query Match 16.0%; Score 153.5; DB 2; Length 1535;
Best Local Similarity 23.6%; Pred. No. 0.00011;
Matches 47; Conservative 27; Mismatches 68; Indels 57; Gaps 7;

QY	4	VDLRWQDCCECDRTTRGQFNA-----PSYHFRGRRLSEFYQEDKPTKTRPKIPSVGRQ	59
DB	1345	INDYLMQEC-----GRGNSPPALFVPTVR-----KPT-PSGDRQKRDGKE	1384
QY	60	GEHLSNSTSAFSTRSDASGTNDFR-----EFVLEMOKTITDRTQIKLE-----	104
DB	1385	NDEVAATRESYDSPLESLYVNERVSGLELLIGFQKEIKLHKIKRIKLEDSGNSADSEP	1444
QY	105	-----SRISTTECVAGSGESHANNTRKWKKDACTICECKDGQVTCVEACP	149
DB	1445	VAQVQLAAPPOLVASKPRSHRCVDDKGTTRLNNEVMSVDVCTCKCFCFGVNCJLBERCG	1504
QY	150	PATC---AVPVNIPGACCP	165
DB	1505	EVSCPPGVPLTAASGGLP	1523

RESULT 2

A39804 thrombospondin precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A39804

R:Lawler, J.; Duguet, M.; Perro, P.

J. Biol. Chem. 266, 8039-8043, 1991

A:Title: Cloning and sequencing of chicken thrombospondin.

A:Reference number: A39804; MID:91217026; PMID:2022631

A:Accession: A39804

A:Status: preliminary

A.Molecule type: mRNA
 A.Residues: 1-1178 <LAA>
 A.Cross-references: GB:M60853; NID:9212763; PIDN:AA4379.1; PID:9212764
 C:Superfamily: thrombospondin 1; EGF homology: thrombospondin type 1 repeat homology; v
 F:325-383/Domain: von Willebrand factor type C repeat homology <VWC>
 F:386-437/Domain: thrombospondin type 1 repeat homology <THR1>
 F:442-498/Domain: thrombospondin type 1 repeat homology <THR2>
 F:499-555/Domain: thrombospondin type 1 repeat homology <THR3>
 F:658-697/Domain: EGF homology <EGF>

Query Match 15.4%; Score 148; DB 1; Length 1178;
 Best Local Similarity 26.0%; Pred. No. 0.00025;

Matches 46; Conservative 24; Mismatches 67; Indels 40; Gaps 8;

QY 28 HFRG-RRSIEFSYQ--ED---KPTKTRPKRIPSVGROGE--HLNSTAFSTRSDAS 77
 ||| :
 Db 207 HRRGLQNHLFDPSIDEVLKKGQSRQSTEVNTINSTELHLS---PAVTFYVGE 263
 QY 78 GTNDFRER-----VLEMQTTIDLRQIKKLESRLSTT-----ECVDAG----- 116
 || :
 Db 264 KTEKAEFCDRSCSELGTWETELTGIRIYVNNLADLQKVESENOIMELIGPNKTLKMQ 323
 QY 117 -----GESHANNTRKMKDACTICECKDQGVTCFEACGPATCAVPNIPGACPPVC 167
 || :
 Db 324 SVCWODGRVFADESSEWVDSCTKCTQDSKIVCHOITCPVSCADPSFEGECPPVC 380

RESULT 3

AS5195

chordin precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A55195

R:Sasali, Y.; Lu, B.; Steinhilber, H.; Gelsert, D.; Gont, L.K.; De Robertis, E.M.
 Cell 79, 779-790, 1994

A>Title: Xenopus chordin: a novel dorsalizing factor activated by organizer-specific hom

A:Reference number: A55195; MUID:95094250; PMID:8001117

A:Accession: A55195

A:Status: preliminary

A.Molecule type: mRNA

A:Residues: 1-941 <SAS>

A.Cross-references: GB:L35764; NID:9603944; PIDN:AA42222.1; PID:9603945

C:Superfamily: chordin; von Willebrand factor type C repeat homology
 F:42-119/Domain: von Willebrand factor type C repeat homology <VWC1>
 F:692-753/Domain: von Willebrand factor type C repeat homology <VWC2>
 F:770-840/Domain: von Willebrand factor type C repeat homology <VWC3>
 F:858-922/Domain: von Willebrand factor type C repeat homology <VWC4>

Query Match 15.3%; Score 147; DB 1; Length 941;
 Best Local Similarity 29.5%; Pred. No. 0.00025;

Matches 43; Conservative 20; Mismatches 45; Indels 38; Gaps 8;

QY 47 KTRPKRIPSVGROGE-----HLNSTAFSTRSDASGTNDFREVLDM-----OKTTIDL 96
 || :
 Db 639 KINPR-----GEIRGOIHIPNCS-----GVSILPEEPPEYELEIEGRQRPDDI 686
 QY 97 RTOIKLESRLSTTECVADGESHANNTRKMKD---ACTICECKDQGVTCFEACGPATC 153
 || :
 Db 687 RK-----DPRACSE-----GOLRAHGRWADPDYDRKSCVSCQKRTVICDPIVCP LNC 736
 QY 154 AVPNIPGACPPVCLOK---RAEKP 176
 || :
 Db 737 SQPVHLPDCCPCVCEKKEKREKRP 762

RESULT 4

TSRHP2

thrombospondin 2 precursor - human

C:Species: Homo sapiens (man)

C:Date: 19-May-1995 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999

C:Accession: A47379; A42173

R:Labell, T.L.; Byers, P.H.
 Genomics 17, 225-229, 1993

A>Title: Sequence and characterization of the complete human thrombospondin 2 cDNA: p

A:Reference number: A47379; MUID:94010892; PMID:8406456

A:Accession: A47379

A.Molecule type: mRNA

A.Residues: 1-1172 <LAA>
 A.Cross-references: GB:L12350; NID:9307505; PIDN:AAA03703.1; PID:9307506

R:Labell, T.L.; Milewicz, D.J.; Distèche, C.M.; Byers, P.H.
 Genomics 12, 421-429, 1992

A>Title: Thrombospondin II: partial cDNA sequence, chromosome location, and expressio

A:Reference number: A42173; MUID:92217961; PMID:1559694

A:Accession: A42173

A.Molecule type: mRNA

A.Residues: 560-1172 <LAA2>

A.Cross-references: GB:M81339

A:Experimental source: fibroblast

A>Note: sequence extracted from NCBI backbone (NCBIN:95091, NCBI:95096)

C:Genetics:

A:Gene: GDB:TMBS2; TSP2

A.Cross-references: GDB:128789; OMIM:188061

A:Map position: 6q27-6q27

C:Complex: homotrimer, disulfide linked

C:Function:

A:Description: participates in cell migration and adhesion, and in platelet aggregati

C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri

F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-1172/Product: thrombospondin 2 #status predicted <MAT>

F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
 F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
 F:436-492/Domain: thrombospondin type 1 repeat homology <THR2>
 F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>

F:553-588/Domain: EGF homology <EGF1>
 F:652-691/Domain: EGF homology <EGF2>

F:928-930/Region: cell attachment (R-G-D) motif
 F:151-316-330-457-584-710-1069/Binding site: carbohydrate (Asn) (covalent) #status pr

F:167-226/Disulfide bonds: interchain #status predicted
 F:266-270/Disulfide bonds: interchain #status predicted

F:612/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 15.3%; Score 147; DB 1; Length 1172;
 Best Local Similarity 48.1%; Pred. No. 0.0003;

Matches 25; Conservative 3; Mismatches 24; Indels 0; Gaps 0;

QY 117 GESHANNTRKMKDACTICECKDQGVTCFEACGPATCAVPNIPGACPPVC 168
 || :
 Db 324 GREFAENETWVDSCTCTCKKFKTICQITCPATCASPSVEGECPSCL 375

RESULT 5

AA3426

collagen alpha 2 fibrillar chain precursor - sea urchin (Strongylocentrotus purpuratu

C:Species: Strongylocentrotus purpuratus (purple urchin)
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999

C:Accession: AA3426
 R:Exposito, J.Y.; D'Alessio, M.; Ramirez, F.
 J. Biol. Chem. 267, 17404-17408, 1992

A>Title: Novel amino-terminal propeptide configuration in a fibrillar procollagen and

A:Reference number: AA3426; MUID:92381062; PMID:1380962
 A:Accession: AA3426
 A:Status: preliminary; not compared with conceptual translation.
 A.Molecule type: nucleic acid
 A.Residues: 1-3198 <EX>
 A.Cross-references: GB:M92041; NID:9161448; PIDN:AAA30040.1; PID:9161449
 A>Note: sequence extracted from NCBI backbone (NCBI:111965)
 C:Superfamily: unassigned collagens; fibrillar collagen carboxyl-terminal homology; v
 F:48-106/Domain: von Willebrand factor type C repeat homology <VWC>
 F:2978-3198/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 14.5%; Score 139.5; DB 2; Length 3198;
 Best Local Similarity 30.9%; Pred. No. 0.0032;

Matches 29; Conservative 15; Mismatches 39; Indels 11; Gaps 2;

QY 83 REFVLEMQTTIDLRQIKKLESRLSTTECVDA-----GESHANNTRKMKDACTI 133

Db 12 RDTLLIATITVFVAVVCGGSESSPFLS--ISSGPELLFCVYRCOPYLHAESWSVDECTI 69
 QY 134 CECKDQVTCFVEACPPATCAVPVNIIPGACCPVC 167
 Db 70 GCGDNPTTTCVIESCPAPFCVPEPIKEGECFLC 103

RESULT 6

thrombospondin 2 precursor - mouse

C:Species: Mus musculus (house mouse)
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
 C:Accession: A42587; A39851
 R:Laberty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seidlin, M.F.; Dixlt, V.M.
 J. Biol. Chem. 267, 3274-3281, 1992
 A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
 A:Reference number: A42587; MID:92147683; PMID:1371115
 A:Accession: A42587
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1172 <IAH>
 A:Cross-references: GB:I07803; GB:M67275; NID:9340421; PIDN:AAA53064.1; PID:9567241
 A:Note: sequence extracted from NCBI backbone (NCBIP:81502)
 R:Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixlt, V.M.
 J. Biol. Chem. 266, 12821-12824, 1991
 A:Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
 A:Reference number: A39851; MID:91302287; PMID:1712771
 A:Accession: A39851
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-873 <BOR>
 A:Cross-references: GB:M64866; NID:9201994; PIDN:AAA40432.1; PID:9201995
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; von
 C:Keywords: calcium binding; glycoprotein
 F:319-377/Domain: von Willebrand factor type C repeat homology <VWC>
 F:380-431/Domain: thrombospondin type 1 repeat homology <THR1>
 F:436-493/Domain: thrombospondin type 1 repeat homology <THR2>
 F:493-549/Domain: thrombospondin type 1 repeat homology <THR3>
 F:553-588/Domain: EGF homology <EGF1>
 F:652-691/Domain: EGF homology <EGF>

Query Match 14.4%; Score 139; DB 2; Length 1172;
 Best Local Similarity 28.6%; Pred. No. 0.0014;
 Matches 30; Conservative 10; Mismatches 43; Indels 22; Gaps 2;

QY 84 EYVLEKQKTTIDLRQIKLESRL-----STTCVDAGESHAN 122
 Db 271 EELSMNMNELSGLHWVNOISKRLERSSDNOFLLELIGPLKTRMSACVOE-GRIFAE 329
 QY 123 NTKMKKDACTCECKDQVTCFVEACPPATCAVPVNIIPGACCPVC 167
 Db 330 NETWVDSCTTCTCKKFKTVCHQITCSPTACANPSTVSGECPC 374

RESULT 7

collagen alpha 5 chain - sea urchin (Strongylocentrotus purpuratus) (fragment)

C:Species: Strongylocentrotus purpuratus (purple urchin)
 C>Date: 20-Jul-1996 #sequence_revision 08-Nov-1996 #text_change 25-Apr-1997
 C:Accession: S63986; S64638
 R:Exposito, J.Y.; Boute, N.; Deleage, G.; Garrone, R.
 Eur. J. Biochem. 234, 59-65, 1995
 A:Title: Characterization of two genes coding for a similar four-cysteine motif of the a
 A:Reference number: S63985; MID:96096722; PMID:8529669
 A:Accession: S63986
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1376 <EXP>
 A:Cross-references: EMBL:X89804
 R:Exposito, J.Y.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: S64637

A:Accession: S64638
 A:Molecule type: DNA
 A:Residues: 1-658, 'G', 660-870, 'G', 872-901, 'H', 903-1185, 'T', 1187-1214, 'Y', 1216-1376 <E
 A:Cross-references: EMBL:X89804
 C:Genetics:
 A:Gene: COL5A1pha
 A:introns: 73/1; 136/2; 221/1; 369/1; 517/1; 659/1; 799/1; 948/1; 1093/1; 1236/1
 A:Superfamily: von Willebrand factor type C repeat homology
 C:Keywords: extracellular matrix
 F:15-73/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 14.2%; Score 137; DB 2; Length 1376;
 Best Local Similarity 41.2%; Pred. No. 0.0023;
 Matches 21; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY 117 GESHANNTKMKDACTICCKDQVTCFVEACPPATCAVPVNIIPGACCPVC 167
 Db 20 GIPYLHEBWKVDECTTCADNATTCVIESCPAPFCVPEPIKEGECFLC 70

RESULT 8

Nei-homolog protein - rat

C:Species: Rattus norvegicus (Norway rat)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
 C:Accession: T10756
 R:Kurda, S.; Tokunaga, C.; Kiyohara, Y.; Konishi, H.; Matsubashi, S.; Kikkawa, U.
 submitted to the EMBL Data Library, November 1998
 A:Description: Protein kinase C-binding protein.
 A:Reference number: Z17122
 A:Accession: T10756
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-810 <KUR>
 A:Cross-references: EMBL:U48246; NID:93851179; PID:93851180
 A:Experimental source: strain Sprague-Dawley, brain

Query Match 14.0%; Score 134.5; DB 2; Length 810;
 Best Local Similarity 32.0%; Pred. No. 0.0022;
 Matches 31; Conservative 14; Mismatches 47; Indels 5; Gaps 3;

QY 78 GTNDFREYVLEKQKTTIDLRQIKLESRLSTTCVDAGESHANNTKN-KKDACTICBC 136
 Db 239 GIMDLOELLAKMTAKLNVAETRLQGLNCHCEKTC-QVSGLLYRDDSDVDDNCGNCTC 297
 QY 137 KDGQVTCFVEACPPATC---AVPVNIIPGACCPVCIQD 170
 Db 298 KSGAVECRRMSPPLNCSPDLSLVHISGCCRCVCRPK 334

RESULT 9

collagen alpha 2 chain precursor - sea urchin (Strongylocentrotus purpuratus) (fragme

C:Species: Strongylocentrotus purpuratus (purple urchin)
 C>Date: 20-Jul-1996 #sequence_revision 01-Nov-1996 #text_change 23-Aug-1997
 C:Accession: S63985; S64637
 R:Exposito, J.Y.; Boute, N.; Deleage, G.; Garrone, R.
 Eur. J. Biochem. 234, 59-65, 1995
 A:Title: Characterization of two genes coding for a similar four-cysteine motif of th
 A:Reference number: S63985; MID:96096722; PMID:8529669
 A:Accession: S63985
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1822 <EXP>
 A:Cross-references: EMBL:X89806
 R:Exposito, J.Y.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: S64637
 A:Accession: S64637
 A:Molecule type: DNA
 A:Residues: 1-381, 'DT', 384-677, 'N', 679-1010, 'L', 1012-1822 <EXM>
 A:Cross-references: EMBL:X89806
 C:Genetics:

A:Gene: COL12a1pha
 A:Introns: 33/1; 106/1; 169/2; 254/1; 395/1; 460/3; 534/1; 678/1; 743/3 817/1; 965/1; 11
 C:Superfamily: unassigned collagens; von Willebrand factor type C repeat homology
 F:48-106/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 14.0%; Score 134.5; DB 2; Length 1822;
 Best Local Similarity 41.0%; Pred. No. 0.0048;
 Matches 25; Conservative 5; Mismatches 20; Indels 11; Gaps 2;

OY 112 CVDAG-----GESHANNTKWKKDACTICECKDGOVTCFEVACPATCAVPNIPGACCPV 166
 DB 49 CVYKGIPIPLHGES-----MSVDECTTCECDNATTTCVIESCQPFCTOPTKPEGCCFL 102

OY 167 C 167
 DB 103 C 103

RESULT 10
 149607
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
 C:Accession: I49607
 R:Amrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
 Dev. Dyn. 195, 113-120, 1992
 A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
 A:Reference number: I49607; MUID:93214071; PMID:1297453
 A:Accession: I49607
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1497 <RES>
 A:Cross-references: GB:I02918; NID:9309180; PIDN:AAA37440.1; PID:9309181
 C:Genetics:
 A:Gene: Col5a-2
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 F:39-98/Domain: von Willebrand factor type C repeat homology <VWC>
 F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 13.8%; Score 132.5; DB 2; Length 1497;
 Best Local Similarity 42.6%; Pred. No. 0.0058;
 Matches 23; Conservative 5; Mismatches 25; Indels 1; Gaps 1;

OY 117 GESHANNTKWKKDACTICECKDGOVTCFEVACPATCAVPNIPGACCPV 169
 DB 44 GOMTLNRDWMKSPCOICVCNCAIICDKIECEVLNCAINDITPGECCPCV 97

RESULT 11
 TSHDPI
 Thrombospondin 1 precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 23-Aug-1987 #sequence_revision 03-Aug-1995 #text_change 17-Nov-2000
 C:Accession: A26155; A34274; A30140; A25812; A05172; A42927
 R:Lavie, J.; Hynes, R.O.
 J. Cell Biol. 103, 1635-1648, 1986
 A:Title: The structure of human thrombospondin, an adhesive glycoprotein with multiple c
 A:Reference number: A26155; MUID:87057617; PMID:2430973
 A:Accession: A26155
 A:Molecule type: mRNA
 A:Residues: 1-1170 <LAW>
 A:Cross-references: GB:X04665; NID:937137; PIDN:CAA28370.1; PID:937138
 A:Note: parts of this sequence, including the amino end of the mature protein, were dete
 R:Laberty, C.D.; Gleaman, T.M.; Dixit, V.M.
 J. Biol. Chem. 264, 11223-11227, 1989
 A:Title: Characterization of the promoter region of the human thrombospondin gene. DNA s
 A:Reference number: A34274; MUID:89291870; PMID:2544587
 A:Accession: A34274
 A:Molecule type: DNA
 A:Residues: 1-166 <LAH>
 A:Cross-references: GB:J04835
 R:Hennessy, S.W.; Frazier, B.A.; Kim, D.D.; Deckwerth, T.L.; Baumgartel, D.M.; Rotwein,
 J. Cell Biol. 108, 729-736, 1989

A:Title: Complete thrombospondin mRNA sequence includes potential regulatory sites in
 A:Reference number: A30140; MUID:89139590; PMID:2918029
 A:Accession: A30140
 A:Molecule type: mRNA
 A:Residues: 1-83, 'A', '85-522, 'A', '524-1170 <HEN>
 A:Cross-references: EMBL:X14787; NID:937464; PIDN:CAA32889.1; PID:937465
 A:Note: parts of this sequence, including the amino end of the mature protein, were d
 R:Kobayashi, S.; Eden-McCutchan, F.; Framson, P.; Bornstein, P.
 Biochemistry 25, 8418-8425, 1986
 A:Title: Partial amino acid sequence of human thrombospondin as determined by analysi
 A:Reference number: A25812; MUID:87157592; PMID:3030396
 A:Accession: A25812
 A:Molecule type: mRNA
 A:Residues: 1-83, 'A', '85-397 <KOB>
 A:Cross-references: GB:M25631; NID:9538353; PIDN:AAA36741.1; PID:9538354
 R:Dixit, V.M.; Hennessy, S.W.; Grant, G.A.; Rotwein, P.; Frazier, W.A.
 Proc. Natl. Acad. Sci. U.S.A. 83, 5449-5453, 1986
 A:Reference number: A05172; MUID:86287276; PMID:3461443
 A:Accession: A05172
 A:Molecule type: mRNA
 A:Residues: 1-83, 'A', '85-374, 'RC' <DIX>
 A:Cross-references: GB:M14326; NID:9340005; PIDN:AAA61237.1; PID:9553801
 A:Note: parts of this sequence, including the amino end of the mature protein, were d
 R:Sun, X.; Skorstengaard, K.; Mosher, D.F.
 J. Cell Biol. 118, 693-701, 1992
 A:Title: Disulfides modulate RGD-inhibitable cell adhesive activity of thrombospondin
 A:Reference number: A42927; MUID:92348511; PMID:1379247
 A:Accession: A42927
 A:Molecule type: protein
 A:Residues: 987-1003 <SUN>
 A:Note: Cys-992 is shown to have a free sulfinydryl
 C:Genetics:
 A:Gene: GDB:THBS1; TSP1; TSP
 A:Cross-references: GDB:120438; OMIM:188060
 A:Map position: 15q15-15q15
 A:Introns: 23/1
 A:Note: the list of introns may be incomplete
 C:Complex: homotrimer, disulfide linked
 C:Function:
 A:Description: participates in cell migration and adhesion, and in platelet aggregati
 C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;
 C:Keywords: beta-hydroxyasparagine; calcium binding; cell adhesion; glycoprotein; tri
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
 F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>
 F:376-429/Domain: thrombospondin type 1 repeat homology <THR1>
 F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
 F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
 F:551-586/Domain: EGF homology <EGF1>
 F:551-586/Domain: EGF homology <EGF1>
 F:551-586/Domain: EGF homology <EGF2>
 F:926-928/Region: cell attachment (R-G-D) motif
 F:171-232/Disulfide bonds: #status predicted
 F:248-360, 708, 1067/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:270, 274/Disulfide bonds: interchain #status predicted
 F:610/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
 F:1051/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 13.0%; Score 125; DB 1; Length 1170;
 Best Local Similarity 29.4%; Pred. No. 0.019;
 Matches 25; Conservative 13; Mismatches 47; Indels 0; Gaps 0;

OY 83 REYLEMOKTTIDRTQIKLKESRLSTTECVADAGESHANNTKWKKDACTICECKDGOV 142
 DB 288 RTIYTTLDGDSIRKYTEENKELANELRRPPLCYHNGOVYRNNEEMTVDSCTECHQONSVTI 347

OY 143 CFVEACPATCAVPNIPGACCPV 167
 DB 348 CKKVCSPIMPSCSNATVDPGECPC 372

RESULT 12
 A40558
 Thrombospondin 1 precursor - mouse

C:\Species: Mus musculus (house mouse)
C:\Date: 05-Jun-1992 #sequence revision 05-Jun-1992 #text_change 20-Aug-1999
C:\Accession: A40558; A37905; B42587; S68787
R:\Lawler, J.; Duquette, M.; Ferro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.
Genomics 11, 587-600, 1991
A:\Title: Characterization of the murine thrombospondin gene.
A:\Reference number: A40558; MUID:92128941; PMID:1774063
A:\Accession: A40558
A:\Status: preliminary
A:\Molecule type: DNA
A:\Residues: 1-1170 <LAW>
A:\Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62455; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:9511867; PIDN:AA50616
J.\Biol.Chem. 265, 16591-16596, 1990
R.\Bornstein, P.; Alfifi, D.; Devaratnam, S.; Fransom, P.; Li, P.
A:\Title: Characterization of the mouse thrombospondin gene and evaluation of the role of
A:\Reference number: A37905; MUID:90375546; PMID:2398070
A:\Accession: A37905
A:\Status: preliminary
A:\Molecule type: DNA
A:\Residues: 1-490 <BOR>
A:\Cross-references: GB:J05605; GB:J05606; NID:q201991; PIDN:AAA0431.1; PID:9554390
J.\Liberty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seiden, M.F.; Dixit, V.M.
J.\Biol.Chem. 267, 3274-3281, 1992
A:\Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:\Reference number: A42587; MUID:92147683; PMID:1371115
A:\Accession: B42587
A:\Status: preliminary; not compared with conceptual translation
A:\Molecule type: mRNA
A:\Residues: 1-1152 'P', 1154-1170 <LAH>
A:\Cross-references: GB:M87276
A:\Note: Sequence extracted from NCBI backbone (NCBI:P.81501)
R.\Chen, H.; Aschlimann, D.; Nowlen, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A:\Title: Expression and initial characterization of recombinant mouse thrombospondin 1 a
A:\Reference number: S68787; MUID:9623406; PMID:8654563
A:\Accession: S68787
A:\Molecule type: protein
A:\Residues: 19-26, 'X', 28-37 <CHE>
C:\Complex: homotrimer, disulfide linked
C:\Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; von
F:\Keywords: calcium binding; glycoprotein; homotrimer
F:\1-170/Product: thrombospondin 1 #status predicted <IG>
F:\1-1170/Product: thrombospondin 1 #status predicted <MAT>
F:\317-375/Domain: von Willebrand factor type C repeat homology <VMC>
F:\378-429/Domain: thrombospondin type 1 repeat homology <THRI>
F:\434-490/Domain: thrombospondin type 1 repeat homology <THR2>
F:\491-547/Domain: thrombospondin type 1 repeat homology <THR3>
F:\551-586/Domain: EGF homology <EGF>
F:\248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.0%; Score 125; DB 2; Length 1170;
Best Local Similarity 29.4%; Pred. No. 0.019;
Matches 25; Conservative 13; Mismatches 47; Indels 0; Gaps 0;

OY 83 REFLVEMQKTTTDRQIKKLSERLSTTECVADGESHSANTKKAKDCTICECKDGQVT 142
Db 288 RTITTYTDDISIRKYTEENRELVSSELKRPLCLFHNGGYGKYKNNEWTVDSCTECHQNSVTI 347

OY 143 CFVEACEPPATCAVPVNIPGACCPYC 167
Db 348 CKRVSCPIMPSCSNATVPDPGCCPCRC 372

RESULT 13
CGHDZY
collagen alpha 2(V) chain precursor - human
C:\Species: Homo sapiens (man)
C:\Date: 31-Jul-1989 #sequence revision 28-Jul-1995 #text_change 31-Dec-2000
C:\Accession: A1147; A25645; SA3643; A25874; I55239; I59025; A25374; A30017
J.\Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J.\Biol.Chem. 264, 2735-2738, 1989
A:\Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the struc-

A:Reference number: A31427; MUID:89123368; PMID:2914927
A:Accession: A31427
A:Molecule type: mRNA
A:Residues: 1-463 <WOO>
A:Cross-references: GB:004478; NID:g179697; PIDN:AAA51859.1; PID:g179698
A:Experimental source: Placenta
R:GreenSPAN, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A:Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for
A:Reference number: A54555; MUID:92314691; PMID:1820205
A:Accession: A54555
A:Molecule type: DNA
A:Residues: 1-32 <GR>
A:Cross-references: GB:M58529; NID:g180834; PIDN:AAC41699.1; PID:g553235
R:Morad-Ameli, M.; Rousseau, J.C.; Klemm, J.P.; Champilaud, M.F.; Boutillon, M.M.;
Eur. J. Biochem. 221, 987-995, 1994
A:Title: Diversity in the processing events at the N-terminus of type-V collagen.
A:Reference number: S43642; MUID:94237164; PMID:8181482
A:Accession: S43643
A:Molecule type: Protein
A:Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>
R:Well, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A:Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibro
A:Reference number: A25874; MUID:87140331; PMID:3029669
A:Accession: A25874
A:Molecule type: mRNA: DNA
A:Residues: 398-1496 <WEI>
A:Cross-references: GB:X04758; PIDN:CAA28454.1; PID:g1340175
A:Experimental source: rhabdomyosarcoma cell line
R:Myers, J.C.; Loidl, H.R.; Stollé, C.A.; Seyer, J.M.
J. Biol. Chem. 260, 5533-5541, 1985
A:Title: Partial covalent structure of the human alpha 2 type V collagen chain.
A:Reference number: I55239; MUID:85182703; PMID:2985598
A:Accession: I55239
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1002-1226 <RE2>
A:Cross-references: GB:M10956; NID:g180427; PIDN:AAA52007.1; PID:g180428
A:Note: Part of this sequence were determined by protein sequencing
R:Emanuel, B.S.; Canizazro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long
A:Reference number: I59025; MUID:85216505; PMID:3858826
A:Accession: I59025
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1003-1034 <RES>
A:Cross-references: GB:M11135; NID:g179693; PIDN:AAA51857.1; PID:g179694
A:Note: Part of this sequence were determined by protein sequencing
R:Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
J. Biol. Chem. 260, 11216-11222, 1985
A:Title: Complete primary structure of the human alpha-2 type V procollagen COOH-termin
A:Reference number: A25374; MUID:85289337; PMID:2411731
A:Accession: A25374
A:Molecule type: mRNA
A:Residues: 1227-1417, 'T', 1419-1437, 'S', 1439-1496 <MEY>
A:Cross-references: GB:M11718; NID:g180912; PIDN:AAA52058.1; PID:g180913
A:Experimental source: normal fibroblasts
R:TSPouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeid, C.S.; Weil, D.; Ramirez, F.
Genomics 3, 275-277, 1988
A:Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located
A:Reference number: A30017; MUID:89138450; PMID:3224983
A:Accession: A30017
A:Molecule type: DNA
A:Residues: 1449-1463, 'E', 1465-1495, 'A' <UST>
A:Cross-references: GB:J03051; NID:g179695; PIDN:AAA51858.1; PID:g179696
A:Note: The authors translated the codon GAA for residue 1460 as Gln, and GAG for res
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
are 5-hydroxylated and subsequently O-glycosylated.
C:Comment: The amino-terminal propeptide domain appears not to be completely cleaved
C:Genetics:
A:Gene: GDB:COL5A2

A:Cross-references: GDB:119064; OMIM:120190
 A:Map position: 2q31-2q31
 A:Introns: 33/1; 812/3; 830/3; 848/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3
 C:Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CHUIV), a
 alpha 2(V) chain and one alpha 3(V) chain. Initially linked by disulfide bonds among the
 length, is formed with desmosine cross-links made from lysine and allysine residues
 C:Function:
 A:Description: structural component of extracellular fibrous polymer associated with cell
 A:Note: may play a role in controlling the lateral growth of collagen I fibrils
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>
 F:27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NP>
 F:27-108/Region: nonhelical
 F:40-99/Domain: von Willebrand factor type C repeat homology <VWC>
 F:109-186/Region: helical
 F:187-208/Region: nonhelical
 F:209-1225/Region: helical
 F:503-505/Region: cell attachment (R-G-D) motif
 F:941-943/Region: cell attachment (R-G-D) motif
 F:1064-1066/Region: cell attachment (R-G-D) motif
 F:1067-1069/Region: cell attachment (R-G-D) motif
 F:1097-1099/Region: cell attachment (R-G-D) motif
 F:1124-1126/Region: cell attachment (R-G-D) motif
 F:1125-1135/Region: cell attachment (R-G-D) motif
 F:1251-1250/Region: carboxyl-terminal nonhelical telopeptide
 F:1251-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F:27/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
 F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:201/Modified site: allysine (Lys) #status predicted
 F:290-293,296,608,614,1004,1007,1013,1028,1034/Modified site: 4-hydroxyproline (Pro) #st
 F:299,1139/Binding site: 5-hydroxylysine (Lys) #status predicted
 F:299,1139/Binding site: carbonyllysine (Lys) (covalent) #status predicted
 F:1025/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:1250-1251/Cleavage site: Gln-Asp (procollagen C-endopeptidase) #status predicted
 F:1250-1251/Binding site: carbonyllysine (Asn) (covalent) #status predicted
 F:1293,1299,1325/Disulfide bonds: interchain #status predicted
 F:1333-1494,1402-1447/Disulfide bonds: #status predicted

Query Match 12.8%; Score 123.5; DB 1; Length 1496;
 Best Local Similarity 42.6%; Pred. No. 0.032;
 Matches 23; Conservative 5; Mismatches 25; Indels 1; Gaps 1;

QY 117 GESNANNTKWKACTICECKRGQVTC-FVEACPATCAVPVNIACCPVCLQ 169
 DB 45 GQMTNRIWDIWPAPCOICVDMGAILDKRICEQDVLDCADPVTPECCPCVCSQ 98

RESULT 14
 S18250
 collagen alpha 1(II) chain precursor - chicken (fragment)
 C:Species: Gallus gallus (chicken)
 C:Date: 22-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
 C:Accession: S18250
 R:Nah, H.D.; Upholt, W.B.
 J. Biol. Chem. 266, 23446-23452, 1991
 A:Title: Type II collagen mRNA containing an alternatively spliced exon predominates in
 A:Reference number: S18250; MUID:92078223; PMID:1744138
 A:Molecule type: mRNA
 A:Residues: 1-206 <NAH>
 A:Cross-references: EMBL:M74435; NID:g211635; PIDN:AAA48714.1; PID:g211636
 C:Genetics:
 A:Gene: COL2A1
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-184/Domain: amino-terminal propeptide #status predicted <PRO>
 F:36-95/Domain: von Willebrand factor type C repeat homology <VWC>
 F:185-206/Product: collagen alpha 1(II) chain (fragment) #status predicted <MAT>

Query Match 11.7%; Score 113; DB 2; Length 206;
 Best Local Similarity 43.1%; Pred. No. 0.035;
 Matches 25; Conservative 5; Mismatches 24; Indels 4; Gaps 4;

QY 112 CYDAGESHANTTKKAKTICECKDQVCFVAC-PPATCAVPNIIP-GACCPVC 167
 DB 37 CVO-DGQSYSDKDWKPPPCRICVDTGTALCEIICEPDCCSP-BIPGECCPVC 92

RESULT 15
 M40333
 collagen alpha 1'(II) chain precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
 C:Accession: M40333
 R:Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solush, M.; Ramirez, F.
 J. Cell Biol. 115, 565-575, 1991
 A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis
 A:Reference number: M40333; MUID:92011898; PMID:1918153
 A:Accession: M40333
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1492 <SUA>
 A:Cross-references: GB:M63596
 A:Note: this sequence is presented as substitutions relative to another sequence in a
 es they replace; the appropriate interpretation of the sequence figure was reconstruct
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F:37-96/Domain: von Willebrand factor type C repeat homology <VWC>
 F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 11.5%; Score 111; DB 2; Length 1492;
 Best Local Similarity 35.4%; Pred. No. 0.33;
 Matches 23; Conservative 7; Mismatches 33; Indels 2; Gaps 2;

QY 104 ESRSTTECVADGESHANNTKWKACTICECKRGQVTCFVEACPATCAVPNIIP-GA 162
 DB 30 EVDLDTGSCVQ-DGORYSPDKDWKPEPCOICVDTGTALCEIICECKDCPNAPBIPGE 88

QY 163 CCPVC 167
 DB 89 CCPVC 93

Search completed: July 24, 2003, 12:53:41
 Job time : 9.6358 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:52:34 ; Search time 2.18936 seconds
(without alignments)
330.094 Million cell updates/sec

Title: US-09-884-319a-2

Perfect score: 962
Sequence: 1 IPRVDLRVWODCCEDCRTRRG.....VNIPGACCPVCIQKRAEKP 176

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 41799 seqs, 4106219 residues

Total number of hits satisfying chosen parameters: 41799

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCF_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	962	100.0	1496	6 US-10-331-496a-28	Sequence 28, Appl
2	123.5	12.8	915	6 US-10-450-186-41	Sequence 41, Appl
3	115	12.0	429	1 PCT-US00-28803-4	Sequence 4, Appl
4	115	12.0	451	1 PCT-US00-28803-2	Sequence 2, Appl
5	107	11.1	1464	1 PCT-US02-18638A-36	Sequence 36, Appl
6	105	10.9	141	6 US-10-249-686-2	Sequence 2, Appl
7	87	9.0	2813	6 US-10-273-573-5521	Sequence 5521, Ap
8	82.5	8.6	354	6 US-10-464-368-58	Sequence 58, Appl
9	82	8.5	357	6 US-10-464-368-59	Sequence 59, Appl
10	81	8.4	348	6 US-10-464-368-56	Sequence 56, Appl
11	81	8.4	348	6 US-10-464-368-54	Sequence 54, Appl
12	79.5	8.3	995	6 US-10-451-901-2	Sequence 2, Appl
13	78.5	8.2	349	6 US-10-464-368-57	Sequence 57, Appl
14	77	8.0	349	1 PCT-US03-12521-2	Sequence 2, Appl
15	77	8.0	349	6 US-10-464-368-55	Sequence 46, Appl
16	77	8.0	1255	6 US-10-451-689-1	Sequence 55, Appl
17	74.5	7.7	1255	6 US-10-451-689-3	Sequence 1, Appl
18	74.5	7.7	1255	6 US-10-273-573-5790	Sequence 3, Appl
19	71.5	7.4	341	6 US-10-273-573-9184	Sequence 5790, Ap
20	71.5	7.4	341	6 US-10-273-573-9184	Sequence 9184, Ap
21	71.5	7.4	347	6 US-10-273-573-8705	Sequence 6705, Ap
22	70.5	7.3	347	6 US-10-273-573-8767	Sequence 8767, Ap
23	70.5	7.3	375	6 US-10-464-368-63	Sequence 63, Appl
24	70	7.3	692	6 US-10-273-573-6812	Sequence 6812, Ap
25	68	7.1	394	6 US-10-273-573-7259	Sequence 7259, Ap
26	68	7.1	1200	6 US-10-460-614-3	Sequence 3, Appl

27	67.5	7.0	819	6 US-10-343-063a-23	Sequence 23, Appl
28	67	7.0	131	6 US-10-273-573-5752	Sequence 5752, Ap
29	67	7.0	574	6 US-10-411-910A-6	Sequence 6, Appl
30	66	6.9	364	6 US-10-451-901-10	Sequence 10, Appl
31	66	6.9	787	5 US-09-291-417D-151	Sequence 151, Appl
32	66	6.9	1413	6 US-10-287-971-389	Sequence 389, App
33	65.5	6.8	517	7 US-60-479-073-287	Sequence 287, App
34	65.5	6.8	1683	7 US-60-470-920-38	Sequence 38, Appl
35	65.5	6.8	1709	6 US-10-370-481-35	Sequence 35, Appl
36	65	6.8	450	6 US-10-411-910A-76	Sequence 76, Appl
37	65	6.8	837	6 US-10-464-368-95	Sequence 95, Appl
38	65	6.8	872	6 US-10-273-573-10007	Sequence 10007, A
39	65	6.8	882	6 US-10-273-573-10011	Sequence 10011, A
40	65	6.8	1074	6 US-10-273-573-10235	Sequence 10235, A
41	65	6.8	1418	6 US-10-273-573-10236	Sequence 10236, A
42	65	6.8	1611	6 US-10-464-368-81	Sequence 81, Appl
43	65	6.8	1615	6 US-10-464-368-82	Sequence 82, Appl
44	65	6.8	1615	6 US-10-374-979-3	Sequence 3, Appl
45	65	6.8	1615	6 US-10-374-979-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-10-331-496a-28
; Sequence 28, Application US/10331496A
; GENERAL INFORMATION:
; APPLICANT: FRANTZ, GRETCHEN
; APPLICANT: HILLAN, KENNETH J.
; APPLICANT: PHILLIPS, HEIDI S.
; APPLICANT: POLAKIS, PAUL
; APPLICANT: SMITH, VICTORIA
; APPLICANT: SPENCER, SUSAN D.
; APPLICANT: WILLIAMS, P. MICKEY
; APPLICANT: WU, THOMAS D.
; APPLICANT: ZHANG, ZEMIN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: P5014R1-PCT
; CURRENT APPLICATION NUMBER: US/10/331,496A
; PRIOR FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US 60/345,444
; PRIOR FILING DATE: 2002-01-02
; PRIOR APPLICATION NUMBER: US 60/351,885
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 60/360,066
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: US 60/362,004
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/366,869
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US 60/366,284
; PRIOR FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 60/368,679
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/404,809
; PRIOR FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: US 60/405,645
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 28
; LENGTH: 1496
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-331-496a-28

Query Match      100.0%; Score 962; DB 6; Length 1496;
Best Local Similarity 100.0%; Pred. No. 7.9e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 IPRVDLRVWODCCEDCRTRGQFNAFSYHFRGRSLSEYQEDKPKRKTPRKIPSYGROG 60
|||||
```

Db 1321 IPRVDLRVMDCCEDCCTRGOFNFAFSYHFRGRSLSEFSYQEDKPTKTRPRKIPSVGRQG 1380
OY 61 EHLNSTAFSTRSDASGTNDFREVLDMOKTITDLPQIKKLESLSTTECVADGESH 120
Db 1381 EHLNSTAFSTRSDASGTNDFREVLDMOKTITDLPQIKKLESLSTTECVADGESH 1440
OY 121 ANNTWKKDDACTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCLQKRAEKP 176
Db 1441 ANNTWKKDDACTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCLQKRAEKP 1496

RESULT 2
US-10-450-186-41

; Sequence 41, Application US/10450186
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; GRIFFIN, Jennifer A.;
; APPLICANT: YAO, Monique G.; DUGGAN, Brendan M.;
; APPLICANT: YUE, Henry; DING, Li;
; APPLICANT: LAU, Preeti G.; LEE, Ernestine A.;
; APPLICANT: RAMKUMAR, Jayalaxmi; THANGAVELU, Kavitha;
; APPLICANT: XU, Yuming; LEE, Sally;
; APPLICANT: TANG, Y. Tom; NGUYEN, Daniel B.;
; APPLICANT: WARREN, Bridget A.; HONCHELL, Cynthia D.;
; APPLICANT: GANDHI, Ameena R.; AKVIZU, Mariah R.;
; APPLICANT: CHAMLA, Nalinder K.; LU, Yan;
; APPLICANT: ELIOTT, Vicki S.; LU, Dzung; Aina M.;
; APPLICANT: HAFALIA, April J A.; AZIMZAI, Yalda;
; APPLICANT: KHAN, Farrah A.; TRAN, Uyen K.
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PI-0345 USN
; CURRENT APPLICATION NUMBER: US/10/450,186
; PRIOR FILING DATE: 2003-06-09
; PRIOR APPLICATION NUMBER: PCT/US01/48517
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255,639
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/257,852
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/260,105
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/262,932
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/263,096
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 60/263,090
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 60/265,926
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PERL Program
; SEQ ID NO 41
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7486326CD1
US-10-450-186-41

Query Match 12.8%; Score 123.5; DB 6; Length 915;
Best Local Similarity 34.5%; Pred. No. 5.6e-05;
Matches 20; Conservative 10; Mismatches 25; Indels 3; Gaps 1;

OY 117 GESNHNNTKKKD---ACTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCLQKRAE 171
Db 669 GQORPHGARMAPNDPLCTSLCTCQRRFVTCDPVCPSPSCPHPVQAPDQCCPVCPEKQ 726

RESULT 3
PCT-US00-28803-4
; Sequence 4, Application PC/TUS0028803
; GENERAL INFORMATION:

; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Donoho, Gregory
; APPLICANT: Nehls, Michael
; APPLICANT: Hilbun, Erin
; APPLICANT: Zambrowicz, Brian
; TITLE OF INVENTION: Novel Human Proteins and Polynucleotides
; FILE REFERENCE: LEX-0070-PCT
; CURRENT APPLICATION NUMBER: PCT/US00/28803
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 60/160,106
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: US 60/162,547
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 429
; TYPE: PRT
; ORGANISM: homo sapiens
PCT-US00-28803-4

Query Match 12.0%; Score 115; DB 1; Length 429;
Best Local Similarity 34.1%; Pred. No. 0.00016;
Matches 15; Conservative 10; Mismatches 19; Indels 0; Gaps 0;

OY 131 CTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCLQKRAE 174
Db 138 CVLCSCTBQIYCGLTTCBPCCPAPLPDPSCCQACADDEASEQ 181

RESULT 4
PCT-US00-28803-2
; Sequence 2, Application PC/TUS0028803
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Donoho, Gregory
; APPLICANT: Nehls, Michael
; APPLICANT: Hilbun, Erin
; APPLICANT: Zambrowicz, Brian
; TITLE OF INVENTION: Novel Human Proteins and Polynucleotides
; FILE REFERENCE: LEX-0070-PCT
; CURRENT APPLICATION NUMBER: PCT/US00/28803
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: US 60/160,106
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: US 60/162,547
; PRIOR FILING DATE: 1999-10-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: homo sapiens
PCT-US00-28803-2

Query Match 12.0%; Score 115; DB 1; Length 451;
Best Local Similarity 34.1%; Pred. No. 0.00017;
Matches 15; Conservative 10; Mismatches 19; Indels 0; Gaps 0;

OY 131 CTICECKDGQVTCFEVACPATCAVPVNIIPGACCPVCLQKRAE 174
Db 138 CVLCSCTBQIYCGLTTCBPCCPAPLPDPSCCQACADDEASEQ 181

RESULT 5
PCT-US02-18638A-36
; Sequence 36, Application PC/TUS0218638A
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc. et al.

;; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
;; FILE REFERENCE: MRI-035PC
;; CURRENT APPLICATION NUMBER: PCT/US02/18638A
;; PRIOR FILING DATE: 2002-06-12
;; PRIOR APPLICATION NUMBER: US 60/298,159
;; PRIOR FILING DATE: 2001-06-13
;; PRIOR APPLICATION NUMBER: US 60/298,155
;; PRIOR FILING DATE: 2001-06-13
;; PRIOR APPLICATION NUMBER: US 60/335,936
;; PRIOR FILING DATE: 2001-11-14
;; NUMBER OF SEQ ID NOS: 238
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 36
;; LENGTH: 1464
;; TYPE: PRT
;; ORGANISM: Homo sapiens
PCT-US02-18638A-36

Query Match
Best Local Similarity 11.1%; Score 107; DB 1; Length 1464;
Matches 24; Conservative 8; Mismatches 31; Indels 12; Gaps 3;

QY 99 QIKLESRLSTECVDAGSGESHANNNTKMKDCTICECKDGQVTCFVEACPPATCAVPVN 158
Db 27 QVGEDEDIPITCYQNGRLRYHDRV-WKPEPCRICVCMNGV-----LCDDVICYDETN 80
QY 159 IPGA-----CCPVC 167
Db 81 CPGAEPPEGCCPVC 95

RESULT 6
US-10-249-686-2
;; Sequence 2, Application US/10249686
;; GENERAL INFORMATION:
;; APPLICANT: Burchardt, Elmar Reinhold
;; TITLE OF INVENTION: Procollagen (III) Propeptides and Related Substances for Treating
;; TITLE OF INVENTION: Fibrotic Diseases
;; FILE REFERENCE: ERB-1
;; CURRENT APPLICATION NUMBER: US/10/249,686
;; CURRENT FILING DATE: 2003-04-30
;; PRIOR APPLICATION NUMBER: PCT/EP01/12663
;; PRIOR FILING DATE: 2001-10-31
;; NUMBER OF SEQ ID NOS: 2
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 2
;; LENGTH: 141
;; TYPE: PRT
;; ORGANISM: human N-terminal procollage (III) propeptide (PIIINP)
US-10-249-686-2

Query Match
Best Local Similarity 10.9%; Score 105; DB 6; Length 141;
Matches 25; Conservative 5; Mismatches 29; Indels 6; Gaps 3;

QY 111 ECVDA-----GESHANNNTKMKDCTICECKDGQVTCFVEACPPATCAVP-VNIP-GACC 164
Db 3 EAVEGCGSHLGOSYARDWKPPEQICVCGSGVLCDDIICDDDELDCPNPEIPGEGC 62
QY 165 PVCLQ 169
Db 63 AVCPQ 67

RESULT 7
US-10-273-573-5521
;; Sequence 5521, Application US/10273573
;; GENERAL INFORMATION:
;; APPLICANT: Hyseq, Inc
;; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES
;; FILE REFERENCE: 21272-066

;; CURRENT APPLICATION NUMBER: US/10/273,573
;; CURRENT FILING DATE: 2002-10-18
;; PRIOR APPLICATION NUMBER: 09/522,929
;; PRIOR FILING DATE: 2000-04-18
;; PRIOR APPLICATION NUMBER: 09/770,160
;; PRIOR FILING DATE: 2001-01-26
;; NUMBER OF SEQ ID NOS: 10994
;; SOFTWARE: Custom
;; SEQ ID NO: 5521
;; LENGTH: 2813
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; NAME/KEY: DOMAIN
;; LOCATION: (826)..(875)
;; OTHER INFORMATION: PRECURSOR GLYCOPROTEIN SIGNAL CELL domain identified by
;; OTHER INFORMATION: EMATRIX, accession number PD02576A, p-value=1.000e-40, raw sco
;; FEATURE:
;; NAME/KEY: DOMAIN
;; LOCATION: (35)..(2402)
;; OTHER INFORMATION: von Willebrand factor type D domain identified by Pfam,
;; OTHER INFORMATION: accession name vwd, E-value=1.7e-183, Pfam score of 623.0
US-10-273-573-5521

Query Match
Best Local Similarity 9.0%; Score 87; DB 6; Length 2813;
Matches 25; Conservative 6; Mismatches 28; Indels 32; Gaps 5;

QY 110 TECVDAGSGESHANNNTKMKD--ACTICECKDG-QVTCFVEACPPA---TCAV----- 155
Db 2555 TGICGEDGYQHGLEAMVDPDHCQICQICLSGRKVNCTQPCTAPAPICGICVARLRQ 2614
QY 156 -----PV-----NIPGACCPVCL 168
Db 2615 MADCCPEYENPCPLGKKEENNTGECGRL 2645

RESULT 8
US-10-464-368-58
;; Sequence 58, Application US/10464368
;; GENERAL INFORMATION:
;; APPLICANT: Krumlauf, Robb
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
;; FILE REFERENCE: 40716-1P-017
;; CURRENT APPLICATION NUMBER: US/10/464,368
;; CURRENT FILING DATE: 2003-06-16
;; PRIOR APPLICATION NUMBER: 60/388,970
;; PRIOR FILING DATE: 2002-06-14
;; NUMBER OF SEQ ID NOS: 140
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO: 58
;; LENGTH: 354
;; TYPE: PRT
;; ORGANISM: MOUSE
US-10-464-368-58

Query Match
Best Local Similarity 8.6%; Score 82.5; DB 6; Length 354;
Matches 17; Conservative 7; Mismatches 24; Indels 7; Gaps 2;

QY 117 GESHANNNTKMKDCTICECKDGQVTCFVEA-----CPATCAVP-VNIPGACC 164
Db 108 GVIYRNGEKEFPNCQYFTCRDQIGCLPRCQLDVLPGPDDCPAPKRVAVPECC 162

RESULT 9
US-10-464-368-59
;; Sequence 59, Application US/10464368
;; GENERAL INFORMATION:
;; APPLICANT: Krumlauf, Robb
;; APPLICANT: Ellies, Debra

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
; FILE REFERENCE: 40716-IP-017
; CURRENT APPLICATION NUMBER: US/10/464,368
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/388,970
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 59
; LENGTH: 357
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-464-368-59

Query Match
Best Local Similarity 33.3%; Score 82; DB 6; Length 357;
Matches 22; Conservative 6; Mismatches 18; Indels 20; Gaps 5;

QY 112 CVDAG-----GESHANNNTKMKKDACTICECKDGQVTCFEVAC-----PPATCAVP--VN 158
Db 110 CVFGDVIRSGEKFQPSCKFQ-----CTCRDGOIGC-VPRQDLVLLPEPCAPARKVE 162
QY 159 IPGACC 164
Db 163 VPGEC 168

RESULT 10
US-10-464-368-56
; Sequence 56, Application US/10464368
; GENERAL INFORMATION:
; APPLICANT: Krumlauf, Robb
; APPLICANT: Eljies, Debra
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
; FILE REFERENCE: 40716-IP-017
; CURRENT APPLICATION NUMBER: US/10/464,368
; CURRENT FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/388,970
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56
; LENGTH: 347
; TYPE: PRT
; ORGANISM: RAT
US-10-464-368-56

Query Match
Best Local Similarity 35.2%; Score 81; DB 6; Length 347;
Matches 25; Conservative 5; Mismatches 19; Indels 22; Gaps 6;

QY 112 CVDAG-----GESHANNNTKMKKDACTICECKDGQVTCFEVAC-----PPATCAVP--VN 158
Db 101 CVFGGSVYRSGEFSQSSCKRYO-----CTCLDGAVGC-VPLCSMDVRLPSPDCPPRRVK 153
QY 159 IPGACC--VC 167
Db 154 LPGKCKEWMVC 164

RESULT 11
US-10-464-368-54
; Sequence 54, Application US/10464368
; GENERAL INFORMATION:
; APPLICANT: Krumlauf, Robb
; APPLICANT: Eljies, Debra
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
; FILE REFERENCE: 40716-IP-017
; CURRENT APPLICATION NUMBER: US/10/464,368
; CURRENT FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/388,970
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 140
```

```

; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 54
; LENGTH: 348
; TYPE: PRT
; ORGANISM: MOUSE
US-10-464-368-54

Query Match
Best Local Similarity 35.2%; Score 81; DB 6; Length 348;
Matches 25; Conservative 5; Mismatches 19; Indels 22; Gaps 6;

QY 112 CVDAG-----GESHANNNTKMKKDACTICECKDGQVTCFEVAC-----PPATCAVP--VN 158
Db 102 CVFGGSVYRSGEFSQSSCKRYO-----CTCLDGAVGC-VPLCSMDVRLPSPDCPPRRVK 154
QY 159 IPGACC--VC 167
Db 155 LPGKCKEWMVC 165

RESULT 12
US-10-451-901-2
; Sequence 2, Application US/10451901
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; BAUGHN, Mariah R.;
; APPLICANT: LU, Yan; ARVIZU, Chandra S.;
; APPLICANT: RAMKUMAR, Jayalaxmi; YAO, Monique G.;
; APPLICANT: POLICKY, Jennifer L.; WALIA, Narinder K.;
; APPLICANT: TRIBOULEY, Catherine M.; YUE, Henry;
; APPLICANT: BATRA, Sajeev; DING, Li;
; APPLICANT: LAU, Preeti G.; BOROMSKI, Mark L.;
; APPLICANT: LU, Dyung Alina M.; GANDHI, Ameena R.;
; APPLICANT: GRIFFIN, Jennifer A.; XU, Yumang;
; APPLICANT: AZIMZAI, Yalda; GIERTZEN, Kimberly J.;
; APPLICANT: TANG, Y. Tom; WARREN, Bridget A.;
; APPLICANT: MASON, Patricia M.; BURFORD, Neil;
; APPLICANT: HAFALIA, April J.A.; LEE, Ernestine A.;
; APPLICANT: YANG, Junmling; GORVAD, Ann E.;
; APPLICANT: EMERLING, Brooke M.; MARQUIS, Joseph P.;
; APPLICANT: LEE, Soo Yeun; SWARNAKAR, Anita;
; APPLICANT: REDDY, Roopa M.; JIANG, Xin;
; APPLICANT: JACKSON, Alan A.
; TITLE OF INVENTION: NUCLEIC ACID-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0869 USN
; CURRENT APPLICATION NUMBER: US/10/451,901
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: PCT/US01/50256
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 60/257,714
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: US 60/260,081
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/262,302
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/263,823
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: US 60/266,088
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 60/348,442
; PRIOR FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 995
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID NO: 926296CD1
US-10-451-901-2

Query Match
Best Local Similarity 18.1%; Score 79.5; DB 6; Length 995;
```

```

Matches 27; Conservative 33; Mismatches 44; Indels 45; Gaps 4;
QY 14 EOCRTGCGNATSYHNRGRSLSESYOE-----KPTKTRPKRP 54
Db 13 QDVVLDESSFS-----ELSSASEEDKEDSNAMEPOKKVPRSKOPPEKSKMP 65
QY 55 SVGROEHLNSNSTAFSTRSDASGTNDREFVLEMOK-----TITDRT----- 98
Db 66 RKKNAPQISDSEVYVYVKEELNSSVAINDTALDEKKNLDVQTLTKTKQKCAQPH 125
QY 99 ---QIKLESRLSTECVDAGSGSHANN 124
Db 126 TVRRFKLKEEETSASMLEGSSNSSET 154

RESULT 13
US-10-464-368-57
; Sequence 57, Application US/10464368
; GENERAL INFORMATION:
; APPLICANT: Krumlauf, Robb
; APPLICANT: Elies, Debra
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION
; FILE REFERENCE: 40716-JP-017
; CURRENT APPLICATION NUMBER: US/10/464,368
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/388,970
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 57
; LENGTH: 349
; TYPE: PRT
; ORGANISM: BOVINE
US-10-464-368-57

Query Match
Best Local Similarity 31.8%; Score 78.5; DB 6; Length 349;
Matches 21; Conservative 8; Mismatches 22; Indels 15; Gaps 4;
QY 117 GESHANNTKMKDACTICCKGQVTCFVEAC-----PPATCAVP--VNIRACCPVCL 168
Db 113 GSFOSCKRYQ-----CTCLDGSVGC-VPLCSVDYRLPSPDCPPRRVKLPKCCCEWV 165
QY 169 OKRAEE 174
Db 166 SRDEKE 171

RESULT 14
PCT-US03-12521-2
; Sequence 2, Application PC/TUS0312521
; GENERAL INFORMATION:
; APPLICANT: Alcon, Inc.
; APPLICANT: Fleenor, Debra L
; APPLICANT: Shepard, Allan
; APPLICANT: Jacobson, Nasreen
; APPLICANT: Pang, Iok-Hou
; APPLICANT: Clark, Abbot F.
; TITLE OF INVENTION: Agents which regulate, inhibit, or modulate the activity and/or
; TITLE OF INVENTION: expression of Connective tissue growth factor (CTGF) as a unique
; TITLE OF INVENTION: lower intracellular pressure and treat glaucomatous retinopathies/
; FILE REFERENCE: 2354 WO F
; CURRENT APPLICATION NUMBER: PCT/US03/12521
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 60/376606
; PRIOR FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 349
; TYPE: PRT
; ORGANISM: homo sapiens

```

```

PCT-US03-12521-2
Query Match
Best Local Similarity 32.4%; Score 77; DB 1; Length 349;
Matches 23; Conservative 7; Mismatches 19; Indels 22; Gaps 6;
QY 112 CVDAG-----GESHANNTKMKDACTICCKGQVTCFVEAC-----PPATCAVP--VN 158
Db 103 CIFGTVYRSGESFOSCKRYQ-----CTCLDGSVGC-MPLCSMDYRLPSPDCPPRRVK 155
QY 159 IPGACCP--VC 167
Db 156 LPKCCCEWVC 166

RESULT 15
PCT-US02-18638A-46
; Sequence 46, Application PC/TUS0218638A
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc. et al.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035PC
; CURRENT APPLICATION NUMBER: PCT/US02/18638A
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-18638A-46

Query Match
Best Local Similarity 32.4%; Score 77; DB 1; Length 349;
Matches 23; Conservative 7; Mismatches 19; Indels 22; Gaps 6;
QY 112 CVDAG-----GESHANNTKMKDACTICCKGQVTCFVEAC-----PPATCAVP--VN 158
Db 103 CIFGTVYRSGESFOSCKRYQ-----CTCLDGSVGC-MPLCSMDYRLPSPDCPPRRVK 155
QY 159 IPGACCP--VC 167
Db 156 LPKCCCEWVC 166

Search completed: July 24, 2003, 13:13:04
Job time : 3.18936 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:53:55 ; Search time 11.9198 Seconds
(without alignments)
1753.524 Million cell updates/sec

Title: US-09-884-319A-2
Perfect score: 962
Sequence: 1 IPRVRLRWQDCCECDCECRTRG.....VNIPGACCPVCLQKRAEKP 176

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5
Searched: 451899 segs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	962	100.0	176	9 US-09-840-707A-6	Sequence 6, Appli
2	962	100.0	176	11 US-09-884-319-2	Sequence 2, Appli
3	962	100.0	176	15 US-10-038-557A-6	Sequence 6, Appli
4	163	16.9	1048	15 US-10-152-724A-4	Sequence 4, Appli
5	160	16.6	400	10 US-09-373-967-2	Sequence 2, Appli
6	160	16.6	1036	10 US-09-373-967-4	Sequence 4, Appli
7	160	16.6	1036	11 US-09-887-527-60	Sequence 60, Appli
8	160	16.6	1036	15 US-10-028-072-142	Sequence 142, App
9	160	16.6	1036	15 US-10-121-049-142	Sequence 142, App
10	160	16.6	1036	15 US-10-123-904-142	Sequence 142, App
11	160	16.6	1036	15 US-10-140-470-142	Sequence 142, App
12	160	16.6	1036	15 US-10-175-746-142	Sequence 142, App
13	160	16.6	1036	15 US-10-176-918-142	Sequence 142, App
14	160	16.6	1036	15 US-10-176-921-142	Sequence 142, App
15	160	16.6	1036	15 US-10-137-865-142	Sequence 142, App

16	160	16.6	1036	15 US-10-140-474-142	Sequence 142, App
17	160	16.6	1036	15 US-10-142-431-142	Sequence 142, App
18	160	16.6	1036	15 US-10-143-114-142	Sequence 142, App
19	160	16.6	1036	15 US-10-140-002-142	Sequence 142, App
20	160	16.6	1036	15 US-10-142-419-142	Sequence 142, App
21	160	16.6	1036	15 US-10-123-262-142	Sequence 142, App
22	160	16.6	1036	15 US-10-142-423-142	Sequence 142, App
23	160	16.6	1036	15 US-10-121-050-142	Sequence 142, App
24	160	16.6	1036	15 US-10-141-755-142	Sequence 142, App
25	160	16.6	1036	15 US-10-143-032-142	Sequence 142, App
26	160	16.6	1036	15 US-10-123-108-142	Sequence 142, App
27	160	16.6	1036	15 US-10-123-236-142	Sequence 142, App
28	160	16.6	1036	15 US-10-123-261-142	Sequence 142, App
29	160	16.6	1036	15 US-10-140-921-142	Sequence 142, App
30	160	16.6	1036	15 US-10-140-928-142	Sequence 142, App
31	160	16.6	1036	15 US-10-121-045-142	Sequence 142, App
32	160	16.6	1036	15 US-10-123-292-142	Sequence 142, App
33	160	16.6	1036	15 US-10-123-903-142	Sequence 142, App
34	160	16.6	1036	15 US-10-124-819-142	Sequence 142, App
35	160	16.6	1036	15 US-10-124-822-142	Sequence 142, App
36	160	16.6	1036	15 US-10-140-925-142	Sequence 142, App
37	160	16.6	1036	15 US-10-160-498-142	Sequence 142, App
38	160	16.6	1036	15 US-10-124-824-142	Sequence 142, App
39	160	16.6	1036	15 US-10-127-825A-142	Sequence 142, App
40	160	16.6	1036	15 US-10-127-829A-142	Sequence 142, App
41	160	16.6	1036	15 US-10-127-835A-142	Sequence 142, App
42	160	16.6	1036	15 US-10-127-839A-142	Sequence 142, App
43	160	16.6	1036	15 US-10-127-901A-142	Sequence 142, App
44	160	16.6	1036	15 US-10-128-693A-142	Sequence 142, App
45	160	16.6	1036	15 US-10-131-813A-142	Sequence 142, App

ALIGNMENTS

RESULT 1
US-09-840-707A-6
; Sequence 6, Application US/09840707A
; Patent No. US2002007276A1
GENERAL INFORMATION:
; APPLICANT: Fredeking, Terry M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
; FILE REFERENCE: 24881-301C
; CURRENT APPLICATION NUMBER: US/09/840,707A
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/562,979
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/198,210
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
FEATURE:
; OTHER INFORMATION: IL-1 receptor intracellular ligand protein
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5,817,476
; PATENT FILING DATE: 1995-06-07
; PUBLICATION DATE: 1998-10-06
US-09-840-707A-6

Query Match 100.0%; Score 962; DB 9; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.2e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IPRVRLRWQDCCECDCECRTRGQFNARFGRRLSFSYQEDKPKTKTRPKRPSVGRG 60
DB 1 IPRVRLRWQDCCECDCECRTRGQFNARFGRRLSFSYQEDKPKTKTRPKRPSVGRG 60

QY	61	EHLSNSTSAFSTSDSGSNDPREVLEMOQTITLRIQIKKLSBSTCTCVDAGESH	120
Db	61	EHLSNSTSAFSTSDSGSNDPREVLEMOQTITLRIQIKKLSBSTCTCVDAGESH	120
QY	121	ANNTKKKKKACATCECKDGOVTCFPEACBPACAPVNIIPGACCPYCLQKAAEKRP	176
Db	121	ANNTKKKKKDCATCECKDGOVTCFPEACBPACAPVNIIPGACCPYCLQKAAEKRP	176

RESULT 2

Sequence 2 Application US/09884319
Publication No. US20030124625A1
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Jiang
Graham, James
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
BINDING
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
City: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/884,319
FILING DATE: 18-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/083,516
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/487,942
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 176 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-884-319-2

	Query Match	Best Local Similarity	Matches 176;	Score 962;	Pred. No.	DB 11;	Length 176;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	IPRYDLRVWODCCEDC	RTRGQFNAFSYHGRSRSLFESYOEDKPKTKTPRPIPSVROG	60						
Dd	1	IPRVDELRWQDCCEDC	ERTGRTGQFNAFSYHGRSRSLFESYOEDKPKTKTPRPIPSVROG	60						
QY	61	EHLNSTSAFSTRSDAS	TGNDFREYLEMOKTITDLRTQIKKLBSRLSTTECVADGESH	120						
Dd	61	EHLNSTSAFSTRSDAS	TGNDFREYLEMOKTITDLRTQIKKLBSRLSTTECVADGESH	120						
QY	121	ANNTKKKKACACTCEC	KDGQAVCFEACAPCAVPVINPGACCVCYLCKRAAEKP	176						
Dd	121	ANNNTKKMKDKACTTC	ECKDKGQAVCFEACAPCAVPVINPGACCVCYLCKRAAEKP	176						

RESULT 3
US-10-038-557A-6

RESULT 3

Sequence 6 Application US/10038557A
Publication ID: US20030092684A1
GENERAL INFORMATION:
APPLICANT: Fredeking, Terry M.
APPLICANT: Ignatyev, George M.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
TITLE OF INVENTION: INJECTIONS AND OTHER DISORDERS
FILE REFERENCE: 24881-301D
CURRENT APPLICATION NUMBER: US/10/038,557A
CURRENT FILING DATE: 2002-06-05
PRIOR APPLICATION NUMBER: 09/840,707
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: 09/562,979
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/198,210
PRIOR FILING DATE: 1999-04-27
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 176
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: IL-1 receptor intracellular ligand protein
PUBLICATION INFORMATION:
PATENT DOCUMENT NUMBER: 5,817,476
PATENT FILING DATE: 1995-06-07
PUBLICATION DATE: 1998-10-06
US-10-038-557A-6

Query Match	100.0%;	Score 962;	DB 15;	Length 176;
Best Local Similarity	100.0%;	Pred. No. 1.2e-89;		
Matches 176;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	IPBVDLRWODCEDCEDRTGQGFNAFSYHRRGRSLSEFSYQDCKPTKTRPAPKPSVGRG	60
Db	1	IPBVDLRWODCEDCEDRTGQGFNAFSYHRRGRSLSEFSYQDCKPTKTRPAPKPSVGRG <td>60</td>	60
Qy	61	EHLNSTSAFSTRSDASGTNDREFVLEKQKITTDLRQIKKLESRLSTPCCVADGESH	120
Db	61	EHLNSTSAFSTRSDASGTNDREFVLEKQKITTDLRQIKKLESRLSTPCCVADGESH	120
Qy	121	ANNTKKKKKACITCECKDGOVTCFEYACDPATCAVPVNI PGACPCVLCQKRAEKP	176
Db	121	ANNTKKKKKACITCECKDGOVTCFEYACDPATCAVPVNI PGACPCVLCQKRAEKP	176

RESULT 4

```

1 Sequence 4 Application US/1015272A4
2 Publication No. US20030082714A1
3
4 GENERAL INFORMATION:
5
6 APPLICANT: LITTLE, Melissa
7 APPLICANT: HOLMES, Gregory
8 APPLICANT: KOLE, Gabriel
9 APPLICANT: YAMADA, Toshiya
10 APPLICANT: GEORGAS, Kyle
11 APPLICANT: WILKINSON, Lorine
12
13 TITLE OF INVENTION: NO. US20030082714A1el Nucleic Acid and Polypeptide
14
15 FILE REFERENCE: P23578
16
17 CURRENT APPLICATION NUMBER: US/10/152,724A
18
19 PRIOR FILING DATE: 2002-05-23
20
21 PRIOR APPLICATION NUMBER: Australian App No. US20030082714A1PQ 4348
22
23 PRIOR FILING DATE: 1999-11-26
24
25 NUMBER OF SEQ ID NOS: 24
26
27 SOFTWARE: PatentIn version 3.1
28
29 SEQ ID NO 4
30
31 LENGTH: 1048

```

US-09-373-967-4

; PRIOR APPLICATION NUMBER: 60/049911
 ; PRIOR FILING DATE: 1997-06-18

1	PRIOR APPLICATION NUMBER: 60/056974
2	PRIOR FILING DATE: 1997-08-26
3	PRIOR APPLICATION NUMBER: 60/059113
4	PRIOR FILING DATE: 1997-09-17
5	PRIOR APPLICATION NUMBER: 60/059115
6	PRIOR FILING DATE: 1997-09-17
7	PRIOR APPLICATION NUMBER: 60/059117
8	PRIOR FILING DATE: 1997-09-17
9	PRIOR APPLICATION NUMBER: 60/059122
10	PRIOR FILING DATE: 1997-09-17
11	PRIOR APPLICATION NUMBER: 60/059184
12	PRIOR FILING DATE: 1997-09-17
13	PRIOR APPLICATION NUMBER: 60/059283
14	PRIOR FILING DATE: 1997-09-18
15	PRIOR APPLICATION NUMBER: 60/059352
16	PRIOR FILING DATE: 1997-09-19
17	PRIOR APPLICATION NUMBER: 60/059588
18	PRIOR FILING DATE: 1997-09-19
19	PRIOR APPLICATION NUMBER: 60/059836
20	PRIOR FILING DATE: 1997-09-24
21	PRIOR APPLICATION NUMBER: 60/062250
22	PRIOR FILING DATE: 1997-10-17
23	PRIOR APPLICATION NUMBER: 60/062814
24	PRIOR FILING DATE: 1997-10-24
25	PRIOR APPLICATION NUMBER: 60/062816
26	PRIOR FILING DATE: 1997-10-24
27	PRIOR APPLICATION NUMBER: 60/063045
28	PRIOR FILING DATE: 1997-10-24
29	PRIOR APPLICATION NUMBER: 60/063082
30	PRIOR FILING DATE: 1997-10-31
31	PRIOR APPLICATION NUMBER: 60/063127
32	PRIOR FILING DATE: 1997-10-24
33	PRIOR APPLICATION NUMBER: 60/063327
34	PRIOR FILING DATE: 1997-10-27
35	PRIOR APPLICATION NUMBER: 60/063329
36	PRIOR FILING DATE: 1997-10-27
37	PRIOR APPLICATION NUMBER: 60/063550
38	PRIOR FILING DATE: 1997-10-28
39	PRIOR APPLICATION NUMBER: 60/063561
40	PRIOR FILING DATE: 1997-10-28
41	PRIOR APPLICATION NUMBER: 60/063704
42	PRIOR FILING DATE: 1997-10-29
43	PRIOR APPLICATION NUMBER: 60/063733
44	PRIOR FILING DATE: 1997-10-29
45	PRIOR APPLICATION NUMBER: 60/063735
46	PRIOR FILING DATE: 1997-10-29
47	PRIOR APPLICATION NUMBER: 60/063738
48	PRIOR FILING DATE: 1997-10-29
49	PRIOR APPLICATION NUMBER: 60/063755
50	PRIOR FILING DATE: 1997-10-17
51	PRIOR APPLICATION NUMBER: 60/064248
52	PRIOR FILING DATE: 1997-11-03
53	PRIOR APPLICATION NUMBER: 60/064809
54	PRIOR FILING DATE: 1997-11-07
55	PRIOR APPLICATION NUMBER: 60/065186
56	PRIOR FILING DATE: 1997-11-12
57	PRIOR APPLICATION NUMBER: 60/065846
58	PRIOR FILING DATE: 1997-11-17
59	PRIOR APPLICATION NUMBER: 60/066364
60	PRIOR FILING DATE: 1997-11-21
61	PRIOR APPLICATION NUMBER: 60/066453
62	PRIOR FILING DATE: 1997-11-24
63	PRIOR APPLICATION NUMBER: 60/066511
64	PRIOR FILING DATE: 1997-11-24
65	PRIOR APPLICATION NUMBER: 60/066770
66	PRIOR FILING DATE: 1997-11-24
67	PRIOR APPLICATION NUMBER: 60/069212
68	PRIOR FILING DATE: 1997-12-11
69	PRIOR APPLICATION NUMBER: 60/069278

PRIOR FILING DATE: 1997-12-11	PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11	PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11	PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-16	PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1998-01-23	PRIOR APPLICATION NUMBER: 60/072320
PRIOR FILING DATE: 1998-01-23	PRIOR APPLICATION NUMBER: 60/073612
PRIOR FILING DATE: 1998-02-04	PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09	PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09	PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12	PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20	PRIOR APPLICATION NUMBER: 60/079224
PRIOR FILING DATE: 1998-03-25	PRIOR APPLICATION NUMBER: 60/079633
PRIOR FILING DATE: 1998-02-27	PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27	PRIOR APPLICATION NUMBER: 60/080155
PRIOR FILING DATE: 1998-03-31	PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081655
PRIOR FILING DATE: 1998-04-14	PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24	PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28	PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12	PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085399
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/085697
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/086414
PRIOR FILING DATE: 1998-05-22	PRIOR APPLICATION NUMBER: 60/086430
PRIOR FILING DATE: 1998-05-22	PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28	PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04	PRIOR APPLICATION NUMBER: 60/088730
PRIOR FILING DATE: 1998-06-10	PRIOR APPLICATION NUMBER: 60/088741
PRIOR FILING DATE: 1998-06-10	PRIOR APPLICATION NUMBER: 60/089334

PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090538
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07

Query Match 16.6%; Score 160; DB 15; Length 1036;
Best Local Similarity 43.1%; Pred. No. 1.8e-07;
Matches 22; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

OY 117 GESHANNTKWKACTICECKDGOVTCFVEACPPATCAVNPINIGACCPVC 167
Db 823 GKAYADEERWDDLSCTHCYCLQGOTLCSTVSCPLPCVPEINVEGSCCPMC 873

RESULT 9

US-10-121-049-142
Sequence 142, Application US/10121049
Publication No. US2003002239A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geriltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C17
CURRENT APPLICATION NUMBER: US/10/121, 049
CURRENT FILING DATE: 2002-04-12
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 142
LENGTH: 1036
TYPE: PRT
ORGANISM: Homo Saplen
US-10-121-049-142

Query Match 16.6%; Score 160; DB 15; Length 1036;
Best Local Similarity 43.1%; Pred. No. 1.8e-07;
Matches 22; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

OY 117 GESHANNTKWKACTICECKDGOVTCFVEACPPATCAVNPINIGACCPVC 167
Db 823 GKAYADEERWDDLSCTHCYCLQGOTLCSTVSCPLPCVPEINVEGSCCPMC 873

RESULT 10

US-10-123-904-142
Sequence 142, Application US/10123904
Publication No. US20030022328A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geriltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C54
CURRENT APPLICATION NUMBER: US/10/123, 904
CURRENT FILING DATE: 2002-04-16
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 142
LENGTH: 1036
TYPE: PRT
ORGANISM: Homo Saplen
US-10-123-904-142

Query Match 16.6%; Score 160; DB 15; Length 1036;
Best Local Similarity 43.1%; Pred. No. 1.8e-07;
Matches 22; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

OY 117 GESHANNTKWKACTICECKDGOVTCFVEACPPATCAVNPINIGACCPVC 167
Db 823 GKAYADEERWDDLSCTHCYCLQGOTLCSTVSCPLPCVPEINVEGSCCPMC 873

RESULT 11

US-10-140-470-142
Sequence 142, Application US/10140470
Publication No. US20030022331A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geriltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K

```

      APPLICANT : Wood,William
      TITLE OF INVENTION : SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
      FILE REFERENCE : P330R1C160
      CURRENT APPLICATION NUMBER : US/10/140,470
      PRIOR APPLICATION REMOVED - See Palm or File Wrapper
      SEQ ID NO 142
      LENGTH : 1036
      TYPE : PROT
      ORGANISM : Homo Sapien
US-10-140-470-142

Query Match          16.6% ; Score 160; DB 15; Length 1036;
Best Local Similarity 43.1%; Pred.No.1.8e-07;
Matches   22; Conservative    11; Mismatches    18; Indels     0; Gaps       0

Yy 117 GESHANNTKKMKKDACTICECKDGVTCVEACPPATCAVPVNIIPGACCPCVC 167
|::|| :: || | | | | | | | | | | | | | | | | | | | | | | |
Db 823 GKAYADERMDLDSCTHCYCIGQGTLCSTVSOPPLPCEVIPINVESGCCPMC 873


RESULT 12
US-10-175-746-142
Sequence 142, Application US/10175746
Publication No. US20030027270A1
GENERAL INFORMATION:
APPLICANT : Baker,Kevin P.
APPLICANT : Beresini,Maureen
APPLICANT : Deforge,Laura
APPLICANT : Desnoyers,Luc
APPLICANT : Filvaroff,Ellen
APPLICANT : Gao,Wei-Qiang
APPLICANT : Gerritsen,Mary E.
APPLICANT : Goddard,Audrey
APPLICANT : Godowski,Paul J.
APPLICANT : Gunney,Austin L.
APPLICANT : Sherwood,Steven
APPLICANT : Smith,Victoria A.
APPLICANT : Stewart,Timothy A.
APPLICANT : Tumas,Daniel
APPLICANT : Watanabe,Colin K
APPLICANT : Wood,William
APPLICANT : Zhang,Zhenlin
TITLE OF INVENTION : SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE : ACIDS ENCODING THE SAME
FILE REFERENCE : P330R1C353
CURRENT FILING DATE : 2002-05-19
PRIOR APPLICATION REMOVED - See File Wrapper or Palm
NUMBER OF SEQ ID NOS : 550
SEQ ID NO 142
LENGTH : 1036
TYPE : PROT
ORGANISM : Homo Sapien
US-10-175-746-142

Query Match          16.6% ; Score 160; DB 15; Length 1036;
Best Local Similarity 43.1%; Pred.No.1.8e-07;
Matches   22; Conservative    11; Mismatches    18; Indels     0; Gaps       0

Yy 117 GESHANNTKKMKKDACTICECKDGVTCVEACPPATCAVPVNIIPGACCPCVC 167
|::|| :: || | | | | | | | | | | | | | | | | | | | | | | |
Db 823 GKAYADERMDLDSCTHCYCIGQGTLCSTVSOPPLPCEVIPINVESGCCPMC 873


RESULT 13
US-10-176-918-142
Sequence 142, Application US/10176918
Publication No. US20030027275A1
GENERAL INFORMATION:
```

```

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney/Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C382
CURRENT APPLICATION NUMBER: US/10/176,918
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 142
LENGTH: 1036
TYPE: PRT
ORGANISM: Homo Sapien
US-10-176-918-142

Query Match      16.6% Score 160; DB 15; Length 1036;
Best local Similarity 43.1%; Pred. No. 1.8e-07;
Matches 22; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

Oy      117 GESHHNNKMKWKDACTICECKDGVTCFVEACPPATCAVPYNIPACCPCVC 167
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      823 GKAVADERMULDSCTHCYCIQGGTLCSTVSCPPLCPYPIINVBSCCCPMC 873

RESULT 14
US-10-176-921-142
Sequence 142, Application US/10176921
General Information:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney/Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C288
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 142
LENGTH: 1036
TYPE: PRT
ORGANISM: Homo Sapien
US-10-176-921-142
```

Query Match 16.6%; Score 160; DB 15; Length 1036;
 Best Local Similarity 43.1%; Pred. No. 1.8e-07;
 Matches 22; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

OY 117 GESHANNTWKKDACTICECKDQVTCFEVACPPATCAVPVNIPIGACCPVC 167
 DB 823 GKAYADEERWMDSCTHCYCILOGTLCSTVSCPLPCVPEPINVEGSCCPMC 873

RESULT 15

US-10-137-865-142
 ; Sequence 142, Application US/10137865
 ; Publication No. US20030032155A1
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: Deforge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Guirney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330R1C154
 ; CURRENT APPLICATION NUMBER: US/10/137,865
 ; CURRENT FILING DATE: 2002-05-03
 ; Prior Application removed - See Palm or File Wrapper
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO: 142
 ; LENGTH: 1036
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-137-865-142

Query Match 16.6%; Score 160; DB 15; Length 1036;
 Best Local Similarity 43.1%; Pred. No. 1.8e-07;
 Matches 22; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

OY 117 GESHANNTWKKDACTICECKDQVTCFEVACPPATCAVPVNIPIGACCPVC 167
 DB 823 GKAYADEERWMDSCTHCYCILOGTLCSTVSCPLPCVPEPINVEGSCCPMC 873

Search completed: July 24, 2003, 13:14:50
 Job time : 12.9198 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:48:39 ; Search time 7.78438 Seconds
(without alignments)
956.622 Million cell updates/sec

Title: US-09-884-319A-2

Perfect score: 962
Sequence: 1 IPKVDLRVWQDCEDCFCRTG.....VNIPGACPFVCLQRAEKP 176

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	962	100.0	176	1	US-08-726-525-2
2	962	100.0	176	2	US-08-487-942-2
3	962	100.0	176	2	US-08-726-036A-2
4	962	100.0	176	4	US-09-083-516-2
5	147	15.3	940	2	US-08-938-365-4
6	147	15.3	941	1	US-08-343-760A-2
7	147	15.3	1172	1	US-08-313-288B-19
8	137.5	14.3	810	2	US-08-820-170A-34
9	137.5	14.3	810	3	US-09-055-699-34
10	137.5	14.3	810	3	US-09-273-565-34
11	137.5	14.3	810	4	US-09-565-538-34
12	137.5	14.3	810	4	US-09-661-468-34
13	137.5	14.3	810	4	US-09-976-165-34
14	123.5	13.0	1170	1	US-08-313-288B-20
15	123.5	12.8	954	2	US-08-749-169A-3
16	123.5	12.8	954	2	US-09-130-032A-3
17	114	11.9	235	4	US-09-602-543-5
18	114	11.9	235	4	US-09-602-543-4
19	112	11.6	218	3	US-08-985-526-1
20	112	11.6	441	3	US-08-985-526-3
21	109	11.3	239	5	PCT-US93-01652-1
22	108	11.2	623	3	US-09-029-348-3
23	108	11.2	626	3	US-09-029-348-2
24	107	11.1	160	1	US-08-479-233-11
25	107	11.1	160	5	PCT-US93-00643-11
26	107	11.1	855	2	US-08-938-365-3
27	107	11.1	867	2	US-08-938-365-2

28	107	11.1	1341	3	US-08-963-825-18	Sequence 18, Appl
29	107	11.1	1341	4	US-09-500-811-18	Sequence 18, Appl
30	107	11.1	1341	4	US-09-570-573-18	Sequence 18, Appl
31	107	11.1	1341	4	US-09-548-608-18	Sequence 18, Appl
32	107	11.1	1461	4	US-09-585-887-9	Sequence 9, Appl1
33	107	11.1	1461	4	US-09-289-578-9	Sequence 9, Appl1
34	105.5	11.0	1442	2	US-08-316-650-12	Sequence 12, Appl
35	105.5	11.0	1442	2	PCT-US95-02251-12	Sequence 12, Appl
36	103.5	10.8	816	2	US-08-820-170A-37	Sequence 37, Appl
37	103.5	10.8	816	3	US-09-055-699-37	Sequence 37, Appl
38	103.5	10.8	816	3	US-09-273-565-37	Sequence 37, Appl
39	103.5	10.8	816	4	US-09-565-538-37	Sequence 37, Appl
40	103.5	10.8	816	4	US-09-661-468-37	Sequence 37, Appl
41	103.5	10.8	816	4	US-09-976-165-37	Sequence 37, Appl
42	92	9.6	94	1	US-07-899-535A-1	Sequence 1, Appl1
43	89.5	9.3	49	4	US-09-602-543-7	Sequence 7, Appl1
44	85.5	8.9	228	4	US-09-182-145-19	Sequence 19, Appl
45	85.5	8.9	229	4	US-09-182-145-99	Sequence 99, Appl

ALIGNMENTS

RESULT 1
US-08-726-525-2
; Sequence 2, Application US/08726525
; Patent No. 5788181
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; NUMBER OF INVENTION: BINDING
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,525
; FILING DATE: 07-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,942
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-726-525-2

Query Match 100.0%; Score 962; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 8e-96;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IPKVDLRVWQDCEDCFCRTGCFNAFSYHFGRRSLSEFSYQEDKPKTRPRKIPSYGROG 60

```

DB      1  IPRVLRWQDCECDECRTRGQFNAPSYHGRGRSLFESYOEDKPTKTRPKRIPSVGRG 60
        |||||||
OY      61  EHLNSTSAFSTRSDASGTNDREFVLEMQKTTTDLRTQIKKLESRLSTTECVADAGESH 120
        |||||||
DB      61  EHLNSTSAFSTRSDASGTNDREFVLEMQKTTTDLRTQIKKLESRLSTTECVADAGESH 120
        |||||||
OY      121 ANNTWKWKDACTICECKDGQVTCFVEACPPATCAVPVNI PGACCPVCLQKRAEKP 176
        |||||||
DB      121 ANNTWKWKDACTICECKDGQVTCFVEACPPATCAVPVNI PGACCPVCLQKRAEKP 176
        |||||||

RESULT 2
US-08-487-942-2
; Sequence 2, Application US/08487942
; Patent No. 5817476
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,942
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-942-2

Query Match      100.0%; Score 962; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 8e-96;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

; Sequence 2, Application US/08726036A
; Patent No. 5981482
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,036A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-726-036A-2

Query Match      100.0%; Score 962; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 8e-96;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1  IPRVLRWQDCECDECRTRGQFNAPSYHGRGRSLFESYOEDKPTKTRPKRIPSVGRG 60
        |||||||
DB      1  IPRVLRWQDCECDECRTRGQFNAPSYHGRGRSLFESYOEDKPTKTRPKRIPSVGRG 60
        |||||||
OY      61  EHLNSTSAFSTRSDASGTNDREFVLEMQKTTTDLRTQIKKLESRLSTTECVADAGESH 120
        |||||||
DB      61  EHLNSTSAFSTRSDASGTNDREFVLEMQKTTTDLRTQIKKLESRLSTTECVADAGESH 120
        |||||||
OY      121 ANNTWKWKDACTICECKDGQVTCFVEACPPATCAVPVNI PGACCPVCLQKRAEKP 176
        |||||||
DB      121 ANNTWKWKDACTICECKDGQVTCFVEACPPATCAVPVNI PGACCPVCLQKRAEKP 176
        |||||||

RESULT 4
US-09-083-516-2
; Sequence 2, Application US/09083516
; Patent No. 6300086
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA

```

COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,516
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,942
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 176 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-883-516-2

Query Match 100.0%; Score 962; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 8e-96;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IRVDLRVWQDCCECRGTGQFNATSYHRRGRSLFSTQEDKPTKTRPKIPSGRGG 60
DB 1 IRVDLRVWQDCCECRGTGQFNATSYHRRGRSLFSTQEDKPTKTRPKIPSGRGG 60
QY 61 EHLNSTSAFSTPSDASGTNDPREFVLEMOKTITDRTQIKKLESRLSTTECVADAGESH 120
DB 61 EHLNSTSAFSTPSDASGTNDPREFVLEMOKTITDRTQIKKLESRLSTTECVADAGESH 120
QY 121 ANNTKKKDACCTCECKDQVTCFVACPCATCAVFNIPGACCPVCLQRAEKP 176
DB 121 ANNTKKKDACCTCECKDQVTCFVACPCATCAVFNIPGACCPVCLQRAEKP 176

RESULT 5
US-08-938-365-4
Sequence 4, Application US/08938365
Patent No. 5989909
GENERAL INFORMATION:
APPLICANT: Yang, Pan
TITLE OF INVENTION: HUCHORDIN AND USES THEREOF
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,365
FILING DATE: 26-SEP-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 09404/040001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 940 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-365-4

Query Match 15.3%; Score 147; DB 2; Length 940;
Best Local Similarity 29.5%; Pred. No. 3.6e-07;
Matches 43; Conservative 20; Mismatches 45; Indels 38; Gaps 8;

QY 47 KTRPKIPSVGRGE-----HLSNSTSAFSTRSDASGTNDPREFVLEMO-----QKTTTL 96
DB 638 KLNPR-----GELRGQIHIPNSCES---GGVSLTPEEPYEEYIYEEGRDPDDL 685
QY 97 RTQIKKLESRLSTTECVADAGESHANNTKMKD---ACTICECKDQVTCFVACPCATC 153
DB 686 RK-----DRACGFE-----GQLRAHGSRAPIYDKKCSVCQKRTVICDPIVCPPLNC 735
QY 154 AVFNIPGACCPVCLQK---RAEKP 176
DB 736 SQPVHLPDQCCPVCEKREKREVKKP 761

RESULT 6
US-08-343-760A-2
Sequence 2, Application US/08343760A
Patent No. 5679783
GENERAL INFORMATION:
APPLICANT: De Robertis, Edward M
TITLE OF INVENTION: Tissue Differentiation Affecting
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Majestic, Parsons, Siebert & Hsue
STREET: Four Embarcadero Center, Suite 1450
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,760A
FILING DATE: 22-NOV-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Siebert, J. Suzanne
REGISTRATION NUMBER: 28,758
REFERENCE/DOCKET NUMBER: 3100.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 363-5556
TELEFAX: (415) 362-5418
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-343-760A-2

Query Match 15.3%; Score 147; DB 1; Length 941;
 Best Local Similarity 29.5%; Pred. No. 3,6e-07;
 Matches 43; Conservative 20; Mismatches 45; Indels 38; Gaps 8;

QY 47 KTRPKRISVSGOGE-----HLSNSTSAFSTRSDASGINDREFVLEW-----QRTTDL 96
 DB 639 KLNPR-----GEIRGOIHIPNSCES-----GVSALTPEEPPEYEIEEGRORDDPL 686
 QY 97 RTQIKLESRLSTTCVADGESHANNNTKMKD---ACTICECKDGYTCFEVACPPATC 153
 DB 687 RK-----DPRACSF-----GOLRAHGSRWAPDYDRKSCVSCQKRTVICDPIVCPPLNC 736
 QY 154 AVPNIPGACPPVLOK---RAEKRP 176
 DB 737 SOPVHLPDQCCPVCEKKREKREVKRP 762

RESULT 7

US-08-313-288B-19
 ; Sequence 19, Application US/08313288B
 ; Patent No. 5750502

GENERAL INFORMATION:
 APPLICANT: Jessell, Thomas M. and Avihu Klar
 TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
 TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/313,288B
 FILING DATE: January 5, 1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0526
 TELEX:
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1172 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-313-288B-19

Query Match 15.3%; Score 147; DB 1; Length 1172;
 Best Local Similarity 48.1%; Pred. No. 4,8e-07;
 Matches 25; Conservative 3; Mismatches 24; Indels 0; Gaps 0;

QY 117 GESANNNTKMKKADTICECKDGYTCFEVACPPATCNPVNIIPGACPPVLC 168
 DB 324 GRFEAENETWVSDCTCTCKKFKTICHOITCPATCASPFESEGCPSCL 375

RESULT 8
 US-08-820-170A-34
 ; Sequence 34, Application US/08820170A

Patent No. 5831058
 GENERAL INFORMATION:

APPLICANT: Tsutomu, FUJIWARA
 APPLICANT: Takeshi, WATANABE
 APPLICANT: Masato, HORIE
 APPLICANT: Toyomasa, KATAGIRI
 TITLE OF INVENTION: HUMAN GENE
 NUMBER OF SEQUENCES: 42
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
 STREET: 2100 Pennsylvania Avenue, N.W.
 CITY: Washington
 STATE: D.C.

COUNTRY: United States
 ZIP: 20037-3202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/820,170A
 FILING DATE:

CLASSIFICATION: 536
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 293-7060
 TELEFAX: (202) 293-7860
 TELEX: 6491103
 INFORMATION FOR SEQ ID NO: 34:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 810 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-820-170A-34

Query Match 14.3%; Score 137.5; DB 2; Length 810;
 Best Local Similarity 30.8%; Pred. No. 3,1e-06;
 Matches 32; Conservative 16; Mismatches 43; Indels 13; Gaps 4;

QY 80 NDERFEV---LEMQRTTIDLRQIKLESRLSTE-----CVDAGESHANNNTKW-KKD 129
 DB 231 SDFLSLVGIMDLQELAKMTAKLVAYETRLSOLNCHCEKCTQVSGLLYRQDSWVDD 290
 QY 130 ACTICECKDGYTCFEVACPPATC---AVPNIPGACPPVLOK 170
 DB 291 HCRNCTCKSGAVECRMSPCLNCSPDPLPVHIAQCKVCNPK 334

RESULT 9
 US-09-055-699-34
 ; Sequence 34, Application US/09055699
 ; Patent No. 6005088

GENERAL INFORMATION:
 APPLICANT: Tsutomu, FUJIWARA
 APPLICANT: Takeshi, WATANABE
 APPLICANT: Masato, HORIE
 APPLICANT: Toyomasa, KATAGIRI
 TITLE OF INVENTION: HUMAN GENE
 NUMBER OF SEQUENCES: 42
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
 STREET: 2100 Pennsylvania Avenue, N.W.
 CITY: Washington
 STATE: D.C.

COUNTRY: United States
 ZIP: 20037-3202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

```

1 APPLICATION NUMBER: US/09/055,659
2
3 FILING DATE:
4
5 CLASSIFICATION:
6
7 PRIOR APPLICATION DATA:
8 APPLICATION NUMBER: 08/820,170
9
10 FILING DATE:
11
12 TELECOMMUNICATION INFORMATION:
13
14 TELEPHONE: (202) 293-7060
15
16 TELEFAX: (202) 293-7860
17
18 TELEX: 6491103
19
20 INFORMATION FOR SEQ ID NO: 34:
21
22 SEQUENCE CHARACTERISTICS:
23
24 LENGTH: 810 amino acids
25
26 TYPE: amino acid
27
28 TOPOLOGY: linear
29
30 MOLECULE TYPE: protein
31
32 US-09-055-659-34

```

Query Match	14.3%	Score 137.5;	DB 3,	Length 810;
Best Local Similarity	30.8%	Pred. No. 3, 1e-06;		
Matches	32;	Conservative	16;	Mismatches 43;
				Indels 13;
				Gaps 4

```

OY      80 NDFEEFV---LEMOKITIDLRTOIKLESSTTE-----CVDAGESHANNTKV-KKD 129
        |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      231 SDFSLVGGIMDLDELIAKKTAKIKNYAETRLSOLENCHCEKTCQVAGSLIRPDQSDVWDG 290

```

```

QY      130 ACTICECKDGOVTCFEVACPATC--AVPYNIPGACCPVCLQK 170
          | | | | | | | | | | | | | | | | | | | |
Db      291 HCRNCTCKSGAVECRRMSCPPLNCSPLVHIAQGCKVCRPK 334

```

RESULT 10
US-09-273-565-34
; Sequence 34, Application US/09273565A
; Data+No 5166100

```

GENERAL INFORMATION:
APPLICANT: FUJIMARA, TSUTOMU
APPLICANT: WATANABE, TAKESHI
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIOUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599
CURRENT APPLICATION NUMBER: US/09/773,565A
CURRENT FILING DATE: 1999-03-22
EARLIER APPLICATION NUMBER: 09/055,699
EARLIER FILING DATE: 1998-04-07
EARLIER APPLICATION NUMBER: 08/820,170
EARLIER FILING DATE: 1997-03-19
EARLIER APPLICATION NUMBER: JP 63410/1996
EARLIER FILING DATE: 1996-03-19
EARLIER APPLICATION NUMBER: JP 69163/1997
EARLIER FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 34
LENGTH: 810
TYPE: prt
ORGANISM: Homo sapiens
US-09-273-565-34

```

Query Match	14.3%	Score 137.5	DB 3	Length 810
Best Local Similarity	30.8%	Pred. No. 3.1e-06		
Matches	32	Conservative	43	Indels 13
				Gaps 4

QY 80 NDFREFV---LEMÖKITITDLRTÖIKKLESRLSTTE-----CVDAGGESHANNTKW-KKD 129

Db 231 SDFSLVQIMDLQELLAKMTAKLNVAETRLSÖLENCHCEKTCQVSGLLYRDQDSWVDGD 290

QY 130 ACTICECKDQVTCFEVACPATC--AVPVNIPGACCPVCLQK 170

Db 291 HCRNCTCKSGAVECRMRMSPPLNCSPDLPVHIAGQCCCKVCRPK 334

RESULT 11
US-09-565-538-34
; Sequence 34, Application US/09565538
; Patent No. 6333404

```

; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE MASATO

```

```

; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
;
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIOUITIN-CONUGATING ENZYME
;
; FILE REFERENCE: Q-53599
;
; CURRENT APPLICATION NUMBER: US/09/565 538

```

```

; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 09/273,565
; PRIOR FILING DATE: 1999-03-22
; PRIOR APPLICATION NUMBER: 09/055,690

```

; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: 08/820,170
 ; PRIOR FILING DATE: 1997-03-19
 ; PRIOR APPLICATION NUMBER: JP 63410/7

; PRIOR FILING DATE: 1996-03-19
 ; PRIOR APPLICATION NUMBER: JP 69163/1997
 ; PRIOR FILING DATE: 1997-03-05
 ; NUMBER OF SEQ ID NOS: 95

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 810
; MYDNA DBM

```

ORGANISM: Homo sapiens
US-09-565-538-34

Query Match	14.3%	Score 137.5	DB 4	Length 810
Best Local Similarity	30.8%	Pred No. 3.1e-06		
Matches 32, Conservative	16	Mismatches 43	Indels 13	Gaps 4

```

QY      80 NDFEFV---LEMOQTITDLTQIKKLESKSTTE-----CVDAGESHANNTKW-KKD 129
      :|| | :||: : : : : |||| | | : : : |
Db     231 SDFUSLVQGIINDLQELAKMTAKLNAETRSQLENCHEKTCQVSGILLYRQDSDWDGD 290

```

```

Oy      130 ACTICECKDGVTCFEVEACPATC--AVPVNIIGACCPVLCQ 170
          | | | | | : | | | : | | | | |
Db      291 HCRNCTCKSGAVECRMRMCPPLNCSPDSLPHVIAQGCKKVCPRK 334

```

RESULT 12
US-09-661-468-34

Patent No. 6576189
GENERAL INFORMATION:
INVENTOR: FUJIMARA, TSUTOMU
APPLICANT: FUJIMARA, TSUTOMU
APPLICANT: WATANABE, TAKESHI
APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: Q-53599

; CURRENT FILING DATE: 2000-09-13
 ; PRIOR APPLICATION NUMBER: 09/055,699
 ; PRIOR FILING DATE: 1998-04-07

PRIOR FILING DATE: 1997-03-19
PRIOR APPLICATION NUMBER: JP 63410/1996
PRIOR FILING DATE: 1996-03-19

```
; PRIOR FILING DATE: 1997-03-05
;
; NUMBER OF SEQ ID NOS: 95
;
; SOFTWARE: PatentIn Ver. 2.1
```

LENGTH: 8.10

ORGANISM: *Homo sapiens*

US-09-661-468-34

```

Query Match Similarity      14.3%; Score 137.5; DB 4; Length 810;
Best Local Similarity      30.8%; Pred No. 3.1e-06;
Matches      32; Conservative 16; Mismatches 43; Indels 13; Gaps 4

Oy      80 NDFREFV---LEMQKTIIDLRQIKLKSRLSTTE-----CVDAGGESHANNTW-KKD 129
       :|||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      221 SDFSLVGIMDLQGLLKMTAKLVAFTRLSQLNCHCEKTCQVSGLLYRQDSWSD 290
       | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      130 ACTICECKDGOVTCFEVEACPPATC--AVPNVIPGACCPVLQK 170
       | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      291 HCRNCTCKSGAVECRMSCPLNCSPDSPVPIHIAQCCKVCPRK 334
       | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-08-313-288B-20
; Sequence 20, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

US-09-976-165-34
; Sequence 34, Application US/09976165
; Patent No. 6562947
; GENERAL INFORMATION:
; APPLICANT: FUJIMURA, TSUNOMU
; APPLICANT: MATNABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: 0-53599
; CURRENT APPLICATION NUMBER: US/09/976,165
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/565,538
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 09/055,699
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: 08/820,170
; PRIOR FILING DATE: 1997-03-19
; PRIOR APPLICATION NUMBER: JP 63410/1996
; PRIOR FILING DATE: 1996-03-19
; PRIOR APPLICATION NUMBER: JP 69163/1997
; PRIOR FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 34
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-165-34

Query Match          14.3%; Score 137.5; DB 4; Length 810;
Best Local Similarity 30.8%; Pred. No. 3.1e-06;
Matches      32; Conservative 16; Mismatches 43; Indels 13; Gaps 4

Oy      80 NDFREFV---LEMQKTIIDLRQIKLKSRLSTTE-----CVDAGGESHANNTW-KKD 129
       :|||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      221 SDFSLVGIMDLQGLLKMTAKLVAFTRLSQLNCHCEKTCQVSGLLYRQDSWSD 290
       | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      130 ACTICECKDGOVTCFEVEACPPATC--AVPNVIPGACCPVLQK 170
       | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      291 HCRNCTCKSGAVECRMSCPLNCSPDSPVPIHIAQCCKVCPRK 334
       | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,288B
FILING DATE: January 5, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0526
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1170 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-313-288B-20

Query Match      13.0%; Score 125; DB 1; Length 1170;
Best Local Similarity 29.4%; Pred. No. 0.00011;
Matches 25; Conservative 13; Mismatches 47; Indels 0; Gaps 0

Oy      83 REFLVEMKQITDRLTQIKLIESRSTIECVADGESAHNNKKMKDKCTICEKDGYT 142
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db      288 RTVTYTTDDSLRKVEENKEELANELRRPPLCYHNGVQRNNEEWTVDSCTECHCONSVTI 347
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Oy      143 CFVEACPPATCAVPYNIGACCPCVC 167
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db      348 CKRVSCPIMPSCNATVPDECCPRC 372
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 15
US-08-749-169A-3
Sequence 3, Application US/08749169A
Patent No. 5846770
GENERAL INFORMATION:
APPLICANT: RACIE, Lisa
APPLICANT: LAVALLIE, Edward
APPLICANT: DEROBERTIS, Edward
TITLE OF INVENTION: CHORDIN COMPOSITIONS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.,
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,169A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: GI 5284
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8260
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 3:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-749-169A-3

```

```

Query Match      12.8%; Score 123.5; DB 2; Length 954;
Best Local Similarity 34.5%; Pred. No. 0.00012;
Matches 20; Conservative 10; Mismatches 25; Indels 3; Gaps 1;

```

```

OY      117 GESHANNTKWKD---ACTICECKDQVTCFVEACPPATCAVPVNIIGACCPVCLQKR 171
          | : : : | : : | : | : | : | : | : | : | : | : | : | : | :
Db       709 GQORPHGARWAPNYDPLCSLCTCQRRTVICDPVVCPPSPCHPVQAPDCCPVCPERQ 766

```

Search completed: July 24, 2003, 12:57:59
 Job time : 9.78438 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:27:29 ; Search time 22.6234 Seconds

(without alignments)
1234.824 Million cell updates/sec

Title: US-09-884-319a-2

Perfect score: 962
Sequence: 1 IPRLVLRWQDCEDCERFRG.....VNIPGACPVCLQKRAEKP 176

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_19Jun03:*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	962	100.0	176	18	AA19989
2	962	100.0	176	21	AA19989
3	962	100.0	550	23	AAE03865
4	962	100.0	550	23	ABG64535
5	962	100.0	814	22	AAU23200
6	962	100.0	1496	20	AAW81030
7	962	100.0	1496	21	AAW70469
8	962	100.0	1496	24	ABU03498
9	924.5	96.1	1498	22	ABU1587

10	390.5	40.6	1463	23	AA021660	Human secreted pro
11	186	19.3	1527	22	AB577771	Drosophila melanog
12	163	16.9	1048	22	AA007143	Chicken CRIM1 prot
13	160	16.6	293	24	AB070387	Human adipocyte Se
14	160	16.6	400	21	AAV82775	Human secreted pro
15	160	16.6	400	21	AAV53033	Human secreted pro
16	160	16.6	1036	21	AAV82776	Human secreted pro
17	160	16.6	1036	21	AAV53034	Human secreted pro
18	160	16.6	1036	22	AAU07141	Human secreted pro
19	160	16.6	1036	22	AAU12242	Human secreted pro
20	160	16.6	1036	23	AAE18852	Human secreted pro
21	160	16.6	1036	24	ABU66640	Human secreted pro
22	160	16.6	1036	24	ABU66640	Human secreted pro
23	160	16.6	1036	24	ABU59721	Human secreted pro
24	157	16.3	1037	22	AAU07142	Human secreted pro
25	149	15.5	732	22	ABG61140	Human secreted pro
26	147	15.3	940	24	ABG75648	Human secreted pro
27	147	15.3	941	18	AAW31559	Human secreted pro
28	147	15.3	1172	21	AAU19677	Human secreted pro
29	147	15.3	1172	21	AAU00043	Human secreted pro
30	147	15.3	1172	23	ABP68622	Human secreted pro
31	147	15.3	1172	23	AAE25031	Human secreted pro
32	147	15.3	1172	23	AAU76902	Human secreted pro
33	147	15.3	1172	23	AAU74788	Human secreted pro
34	147	15.3	1172	24	ABG74674	Human secreted pro
35	147	15.3	1172	24	ABG72839	Human secreted pro
36	145.5	15.1	1593	24	ABP97368	Human secreted pro
37	145.5	15.1	1628	24	ABP97368	Human secreted pro
38	145	15.1	1342	24	ABP97379	Human secreted pro
39	145	15.1	1427	23	ABG79171	Human secreted pro
40	145	15.1	1472	24	ABP97371	Human secreted pro
41	145	15.1	1512	24	ABP97372	Human secreted pro
42	145	15.1	1535	24	ABP97374	Human secreted pro
43	145	15.1	1570	24	ABP97373	Human secreted pro
44	140.5	14.6	63	22	ABG49752	Human secreted pro
45	140.5	14.6	63	22	ABG29742	Human secreted pro

ALIGNMENTS

RESULT 1
AAW19989 standard; Protein: 176 AA.
ID AAW19989:
XX
AC AAW19989:
XX
DT 27-AUG-1997 (first entry)
XX
DE Type I, p80 IL-1-receptor intracellular domain ligand.
XX
KW IL-1, interleukin; receptor; ligand; screening assay; inhibitor;
KW IL-1 mediated response; inflammation; inflammatory; antibody;
KW Intracellular domain; CAMP; calcium activated neutral protease.
XX
OS Homo sapiens.
XX
PN W09640907-A1.
XX
PD 19-DEC-1996.
XX
PF 06-MAY-1996; 96WO-US06363.
XX
PR 07-JUN-1995; 95US-0487942.
XX
PA (GEMV) GENETICS INST INC.
XX
PI Graham J, Lin L;
XX
DR WPI, 1997-052315/05.
XX
DR N-PDB; AAT71216.
XX
PT Interleukin-1 receptor intracellular ligand proteins and related DNA

XX WPI: 2001-343795/36.
 DR N-PSDB; AAD08330.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 XX used in preventing, treating or ameliorating a medical condition -
 PS Claim 11; Page 523-525; 553pp; English.

CC AAD08283-AAD08355 represent cDNAs corresponding to 23 human secreted
 CC protein genes, and AAE03818-AAE03870 represent the proteins they encode.
 CC AAE03871-AAE03896 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 23 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention.

XX
 SQ Sequence 550 AA;
 Query Match 100.0%; Score 962; DB 22; Length 550;
 Best Local Similarity 100.0%; Pred. No. 8,5e-85;
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IPRVDLRVWODCCEDCRTRGQFNAFSYHFGRRSLSESYOEDKPTKTRPKRIPSVGRG 60
 Db 375 IPRVDLRVWODCCEDCRTRGQFNAFSYHFGRRSLSESYOEDKPTKTRPKRIPSVGRG 434

QY 61 EHLNSTSAFSTRSDASGTNDREFVLEMOKITDRLTQIKKLESRLSTTECVADAGESH 120
 Db 435 EHLNSTSAFSTRSDASGTNDREFVLEMOKITDRLTQIKKLESRLSTTECVADAGESH 494

QY 121 ANNTKMKKDACCTICECKDGOVTCFEVACPATCAVPVNIIPGACCPVLOKRAEEKP 176
 Db 495 ANNTKMKKDACCTICECKDGOVTCFEVACPATCAVPVNIIPGACCPVLOKRAEEKP 550

RESULT 4
 ABG64535
 ID ABG64535 standard; Protein; 550 AA.
 XX
 AC ABG64535;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE Human albumin fusion protein #1210.
 XX
 KW Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antiinfertility; antiinflammatory; anticancer;

KW immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200177137-A1.
 XX
 PD 18-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-US11988.
 XX
 PR 12-APR-2000; 2000US-229358P.
 PR 25-APR-2000; 2000US-199384P.
 PR 21-DEC-2000; 2000US-256931P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Haseltine WA;
 DR WPI: 2002-010886/01.
 XX
 PT New fusion protein for treating disease e.g. diabetes comprises an
 PT albumin fused to a therapeutic protein -
 XX
 PS Claim 1; Page 1301-1303; 2102pp; English.

XX The present invention relates to albumin fusion proteins comprising a
 CC therapeutic protein X and human albumin (HA, also known as human serum
 CC albumin, HSA). The proteins are useful for treating a disease or
 CC disorder that may be modulated by therapeutic protein X. The albumin
 CC extends the shelf-life of protein X, and may increase its biological
 CC in vitro/in vivo activity. The protein is useful for treating and
 CC diagnosing disorders such as cancer, reproductive disorders, digestive
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders
 CC (e.g. diabetes), haematopoietic disorders, neural disorders
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders
 CC (e.g. osteoporosis, arthritis). ABG63346-ABG65518 represent albumin
 CC fusion proteins of the invention.

XX
 SQ Sequence 550 AA;
 Query Match 100.0%; Score 962; DB 23; Length 550;
 Best Local Similarity 100.0%; Pred. No. 8,5e-85;
 Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IPRVDLRVWODCCEDCRTRGQFNAFSYHFGRRSLSESYOEDKPTKTRPKRIPSVGRG 60
 Db 375 IPRVDLRVWODCCEDCRTRGQFNAFSYHFGRRSLSESYOEDKPTKTRPKRIPSVGRG 434

QY 61 EHLNSTSAFSTRSDASGTNDREFVLEMOKITDRLTQIKKLESRLSTTECVADAGESH 120
 Db 435 EHLNSTSAFSTRSDASGTNDREFVLEMOKITDRLTQIKKLESRLSTTECVADAGESH 494

QY 121 ANNTKMKKDACCTICECKDGOVTCFEVACPATCAVPVNIIPGACCPVLOKRAEEKP 176
 Db 495 ANNTKMKKDACCTICECKDGOVTCFEVACPATCAVPVNIIPGACCPVLOKRAEEKP 550

RESULT 5.
 AAU23200
 ID AAU23200 standard; Protein; 814 AA.
 XX
 AC AAU23200;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human enzyme polypeptide #286.
 XX
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
XX WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001: 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226779.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0244126.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.

	PR	08-DEC-2000;	2000US-0251989.
	PR	08-DEC-2000;	2000US-0251990.
	PT	11-DEC-2000;	2000US-0254097.
	PR	05-JAN-2001;	2001US-0259678.
	XX	(HUMA-) HUMAN GENOME SCI INC.	
	PA	Rosen CA,	Barash SC, Ruben SM;
	PI	WPI:	2001-465566/50.
	DR	N-PDB;	AAS41070.
	XX	Newel polypeptides and poly nucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancers diseases	
	PS	Claim 11; SEQ ID NO 1196;	1180bp; English.
	CC	The present invention relates to the isolation of novel human enzyme polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g., cancer), immunodeficiency disorders (e.g., AIDS) autoimmune disorders (e.g., arthritis), neurological disorders (e.g., Alzheimer's disease), metabolic disorders (e.g., phenylketonuria), inflammatory disorders (e.g., asthma), cardiovascular disorders (e.g., arteriosclerosis), blood-related disorders (e.g., haemophilia), reproductive disorders (e.g., infertility) and infectious disorders (e.g., influenza). The poly nucleotides of the invention can also be used in gene therapy. AAU22915-AAU23814 represent the novel human enzyme polypeptides of the invention.	
	CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
	CC	Sequence	814 AA;
	XX	Query Match	100.0%; Score 962; DB 22; Length 814; Best Local Similarity 100.0%; Pred. NO. 1.4e-84; Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	I	IPRVLDVWDQCCEDCRTRGOFNAFSYHFRRRSLSFYSQBKPKTKRPRIKIPSGRG	60
Db	639	IPRVDLRWMDCCEDCRTRGFNFASYHFGRRLSFEYGDKPKTKRPPRIEVRGG	698
QY	61	EHLNSTNSASTSRDASGTNDRFREFVLDMQTITIDLRLQIKLERSLTTECVADAGGS	120
Db	699	EHLNSTNSASTSTRDASGTNDFREFFLEMQTTITIDLRQIKLESRLSTTECVADAGGS	758
QY	121	ANNNTKKAKKDACTICECKDGQYTCEVEACPATCAVPNIPEACCPVCLOKRAEKRP	176
Db	759	NNNKMKKDDACTIOECKDGQVTCEVEACPATCAVPNIPEACCPVCLOKRAEKRP	814
RESULT 6	AAAWB1030	standard; Protein; 1496 AA.	
ID	AAWB1030	standard; Protein; 1496 AA.	
AC	AAWB1030;		
DT	10-MAY-1999	(first entry)	
XX	Melanoma associated antigen MG50.		
DE	MG50: melanoma gene-50; melanoma associated antigen; human;		
KM	T cell epitope; cancer; lung cancer; rhabdomyosarcoma; diagnosis;		
KN	therapy; vaccine.		
SS	Homo sapiens		

P	N		XN0855133-AI.	
F	D		10-DEC-1998.	
X	X			
X	X		04-JUN-1998;	98WO-US11533.
X	X			
P	R		06-JUN-1997;	97US-0870941.
X	X			
X	X		(REGC) UNIV CALIFORNIA,	
X	X		(UTSC-) UNIV SOUTHERN CALIFORNIA.	
X	X			
P	I		Deans RJ, Kan-Mitchell J, Minev BR, Mitchell MS;	
D	R		WPI: 1999-080820/07.	
X	X		N-PDB: AAV99922.	
P	T		New MG50 melanoma associated antigen fragments - used to develop	
P	T		products for the detection, treatment and prevention of	
P	T		MG50-expressing cancers, e.g. melanoma, lung cancer or	
X	X		rhabdomyosarcoma	
P	S		Claim 1; Page 45-49; 79pp: English.	
X	X			
C	C		This polypeptide comprises a portion of a new human melanoma	
C	C		associated antigen, designated MG50. The amino acid sequence was	
C	C		deduced from a cDNA clone (see AAV99922) isolated from melanoma	
C	C		cell line M5M M-1 cDNA by subtractive hybridisation. The 5' region	
C	C		of MG50 cDNA was not obtained. MG50 mRNA has been detected in	
C	C		melanoma, lung carcinoma and rhabdomyosarcoma cells, foetal brain,	
C	C		foetal heart and human placenta. The invention also provides T	
C	C		cell epitopes (see AA#B1031-54) from MG50, including cytotoxic and	
C	C		helper T cell epitopes, antibodies that specifically bind to MG50	
C	C		or an MG50 T cell epitope, recombinant vectors, and antigen	
C	C		presenting cells. Methods are provided for identifying an MG50	
C	C		melanoma associated antigen in an individual and for identifying an	
C	C		immune response against an MG50 melanoma associated antigen, as	
C	C		well as methods of stimulating T lymphocytes that are reactive	
C	C		against cancer cells expressing MG50 and for treating an individual	
C	C		having cancer cells that express MG50. The products and methods	
C	C		can be used for the detection, treatment and prevention of	
C	C		MG50-expressing cancers, e.g. melanomas, lung cancer or	
C	C		rhabdomyosarcoma.	
X	X			
S	Q		Sequence	1496 AA;
O	U		Query Match	100.0%; Score 962; DB 20; Length 1496;
B	E		Best Local Similarity	100.0%; Pred. No. 3e+84;
M	a		Matches	176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y	Y	1	IPRVLDLHVWDCCEDCDCTRGCFNAFSYHFRGRSLERSYOEDKPTKKTRPKIDSVGROG	60
D	b	1321	IPRVDLTVMWCCEDCDCTRGCFFNFASFHFRRRSRLERSYODBKPTKKTRPRKISVSGOG	1380
Y	Y	61	EHLNSNSTSAFSTRSDASTGNDFRFVLEEMOKTINDLRIOIKLSRLSTFTCVAGGESH	120
D	b	1381	EHLNSNSTSARSTRSDASTGNDFRFVLEEMOKTTINDLRIOIKLSRLSTFTCVAGGESH	1440
Y	Y	121	ANNTKMKKDACTICECKDGQYTCEVEACPPATCAVPNIIGACCPCVLQKRAEERK	176
D	b	1441	ANNTKMKKDACTICECKDGQYTCEVEACPPATCAVPNIIGACCPCVLQKRAEERK	1496
R	E		RESULT 7	
I	D		AAV70469	
X	X		AAV70469 standard; Protein; 1496 AA.	
X	X		AAV70469;	
D	T		21-JUN-2000 (first entry)	
X	X			
D	e		Human p53 target molecule, PRG2 protein.	

OY 121 ANNTKMKKDACITCECKDGOVTCFVEACPPATCAVPVNIPIGACCPVCLQRAEKP 176
 DB 1441 ANNTKMKKDACITCECKDGOVTCFVEACPPATCAVPVNIPIGACCPVCLQRAEKP 1496
 RESULT 9
 ABB11587 standard; peptide; 1498 AA.
 ID ABB11587
 AC ABB11587;
 DE 11-JAN-2002 (first entry)
 XX Human peroxidase homologue, SEQ ID NO:1957.
 XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; anti-inflammatory;
 KW antistatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytosolic; osteopathic; vasotropic; cardiac; virucide; antibacterial;
 KW antifungal; vulnery; antitumor.
 XX Homo sapiens.
 OS MO200157188-A2.
 PN 09-AUG-2001.
 PD 05-FEB-2001; 2001WO-US03800.
 PF 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT;
 PI WPI: 2001-457740/49.
 XX N-PSDB; ABA08831.
 DR Human proteins and DNA encoding sequences useful for preventing,
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer -
 XX Claim 20: Page 216-217; 1963pp; English.
 PS Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent 1350 novel human polypeptides, and
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical

CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g. asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention.
 SQ Sequence 1498 AA:
 Query Match 96.1%; Score 924.5; DB 22; Length 1498;
 Best Local Similarity 97.2%; Pred. No. 1.3e-80;
 Matches 172; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
 OY 1 IPVDLRVWODCECDRTGQFNAPSYHFRGRSLFEFSYOEOKPTKTRPKIPSYRGOG 60
 DB 1322 IPVDLRVWODCECDRTGQFNAPSYHFRGRSLFEFSYOEOKPTKTRPKIPSYRGOG 1381
 OY 61 EHLNSTSAFSTRSDASGTNDF-REFVLEMOKITITDRLQIKKLESRLSTTECVADGGS 119
 DB 1382 EHLNSTSAFSTRSDASGTNDFORVCSWEMOKITITDRLQIKKLESRLSTTECVADGGS 1441
 OY 120 HANNTKMKKDACITCECKDGOVTCFVEACPPATCAVPVNIPIGACCPVCLQRAEKP 176
 DB 1442 HANNTKMKKDACITCECKDGOVTCFVEACPPATCAVPVNIPIGACCPVCLQRAEKP 1498
 RESULT 10
 ID AAO21660 standard; Protein; 1463 AA.
 XX AAO21660;
 AC AAO21660;
 XX 05-SEP-2002 (first entry)
 DT Human secreted protein SEQ ID No. 2.
 XX
 DE Antiartherosclerotic; cytosolic; HIV; antiallergic; antianaemic;
 KW antistatic; cardiac; vasotropic; neuroprotective; nootropic; SECP;
 KW anticonvulsant; antiparkinsonian; cerebroprotective; antiinflammatory;
 KW immunosuppressive; human secreted protein; cell proliferative disorder;
 KW arteriosclerosis; cancer; autoimmune; inflammatory disorder; AIDS;
 KW allergy; anaemia; asthma; cardiovascular disease; developmental disorder;
 KW ischaemic heart disease; congestive heart failure; neurological disorder;
 KW renal tubular acidosis; hypothyroidism; Alzheimer's disease; dementia;
 KW Parkinson's disease; epilepsy; stroke; knockin humanised animal;
 KW transgenic animal; gene therapy.
 OS Homo sapiens.
 PN WO200238602-A2.
 XX 16-MAY-2002.
 PD 08-NOV-2001; 2001WO-US47420.
 XX 08-NOV-2000; 2000US-247505P.
 PF 09-NOV-2000; 2000US-248642P.
 PR 16-NOV-2000; 2000US-249824P.
 PR 21-NOV-2000; 2000US-252824P.
 PR 08-DEC-2000; 2000US-254305P.
 PR 18-DEC-2000; 2000US-256448P.

XX (INCY-) INCYTE GENOMICS INC.
 PA Yue H, Yao MG, Ganthi AR, Baughn MR, Swarnakar A, Walla NK;
 PI Sanjanwala M, Thornton M, Elliott VS, Lu Y, Gietzen KJ, Buford N;
 PI Ding L, Hafalia AJA, Tang YT, Bandman O, Warren BA, Honchell CD;
 PI Lu DAM, Thanagaveilu K, Lee S, Xu Y, Yang J, Lal PG, Tran B;
 PI Ison CH, Duggan BM, Saperstein SK;
 DR MPI: 2002-519296/55.
 DR N-PSDB: AAL39621.
 XX
 PF Human secreted proteins and polynucleotides for diagnosing, treating or
 PT preventing disorders of cell proliferative, cardiovascular,
 PR developmental, neurological and autoimmune/inflammatory disorders -
 PS
 PS Claim 1: Page 150-153; 229pp; English.
 CC The invention relates to an isolated human secreted protein (SECP)
 CC polypeptide from 63 fully defined protein sequences given in the
 CC specification. The polypeptide is useful for the diagnosing/treating of a
 CC disease with decreased/overexpression of SECP. Examples of disorders
 CC associated with abnormal expression of SECP include a cell proliferative
 CC disorder e.g. arteriosclerosis, cancers; autoimmune/inflammatory
 CC disorder, AIDS, allergies, anaemia, asthma; cardiovascular disease e.g.
 CC congestive heart failure, ischaemic heart disease; developmental disorder
 CC e.g. renal tubular acidosis, hypothyroidism; neurological disorder e.g.
 CC Alzheimer's disease, dementia, Parkinson's disease, epilepsy or stroke.
 CC The SECP polynucleotide and polypeptide are further useful for analysing
 CC the proteome of a tissue or a cell type. The polynucleotide is useful for
 CC creating knockin humanised animals (pigs) or transgenic animals (mice or
 CC rats) to model human disease, and for somatic or germline gene therapy,
 CC and further for generating hybridisation probes useful in mapping the
 CC naturally occurring genomic sequence. This sequence represents a human
 CC secreted protein of the invention.
 XX
 SQ Sequence 1463 AA;
 Query Match 40.6%; Score 390.5; DB 23; Length 1463;
 Best Local Similarity 45.7%; Pred. No. 1e-28;
 Matches 79; Conservative 18; Mismatches 61; Indels 15; Gaps 3;
 QY 1 IPVDLRWQDCCECCRTGQFNAFSYHFRGRSLEFSYQEDKPTKTRPKRIPVSGRO- 59
 DB 1287 IPKVDLRWQDCACCRSGQFRAYTOESQKRSAYSPYDKDELHLR-----SRQ 1341
 QY 60 -----GEHLSNSTSAFSTRSDASGNDREFVLEMOKITDRLTOIKKLSRLSTTECDV 114
 DB 1342 DKTIYGEDARNTVLAKEF-----SODSTFAETIOETITLRLQINKLEARKLRQAGCTD 1397
 QY 115 AGGESHANNTKWKKDACTICECKDGOVTCFEACPPATCAVNPVIGACCPVC 167
 DB 1398 VRGVKRAKBERMKEDCTHCICESGOVTCVEICPPAPCPSELVKGCTCPCVC 1450
 RESULT 11
 ABB57771
 ID ABB57771 standard; Protein; 1527 AA.
 XX ABB57771;
 AC
 XX 26-MAR-2002 (first entry)
 DT
 XX Drosophila melanogaster polypeptide SEQ ID NO 105.
 DE
 XX Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD

XX 23-MAR-2001; 2001WO-0509231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PR
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 DR MPI: 2001-656860/75.
 DR N-PSDB: ABL01874.
 XX
 PF New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PR interactions -
 PS
 PS Disclosure: SEQ ID NO 105; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1527 AA;
 Query Match 19.3%; Score 186; DB 22; Length 1527;
 Best Local Similarity 25.9%; Pred. No. 8.3e-09;
 Matches 51; Conservative 27; Mismatches 71; Indels 48; Gaps 6;
 QY 4 VDLRWQDCCECCRTGQFNAFSYHFRGRSLEFSYQEDKPTKTRPKRIPVSGROGHL 63
 DB 1342 INLYWQEC-----GRCSNPPALFD-----SYIPQTYKRRNRQK-RDLGKRENDV 1386
 QY 64 SNSTSAFSTRSDASGNDPFR-----EFVLEMOKITDRLTOIKKLE----- 104
 DB 1387 AVAESYDPSLESYLVYNERRYSGLLELIGSFQKELEKLLKRLKLEDCSNSADSEPVAVQ 1446
 QY 105 -----SRLSTTECDVAGGESHANNTKWKKDACTICECKDGOVTCFEACPPATC 153
 DB 1447 VOLAAPPOLVSKPKRSHCVDDKGTTRLNNEVWSPDVCCKCCKPRGQVWCLRERGGEVSC 1506
 QY 154 --AVPVNIPGACCPVC 167
 DB 1507 PRGVDPFLPPEACCPHC 1523
 RESULT 12
 AAU07143
 ID AAU07143 standard; Protein; 1048 AA.
 XX AAU07143;
 AC
 XX 24-OCT-2001 (first entry)
 DT
 XX Chicken CRIM1 protein.
 DE
 XX CRIM-1; Chicken: human chromosome 2p21-16.3; ophthalmological;
 KW neuroprotective; renal; osteopathic; dental; vulnerrary; immunogen;
 KW antibody; gene therapy; neurodegenerative disease; eye disorder;
 KW cataract; bone morphogenic protein; BWG; renal disease; bone abnormality;
 KW tooth abnormality; wound; s52.
 XX
 OS Gallus gallus.
 XX
 XX Key Location/Qualifiers
 FH

FT	Region	212..219	/note= "Conserved N-terminal motif"
FT	Region	348..402	/label= CR_1
FT	Region	415..468	/note= "Cysteine rich repeat"
FT	Region	415..468	/label= CR_2
FT	Region	620..674	/note= "Cysteine rich repeat"
FT	Region	620..674	/label= CR_3
FT	Region	691..746	/note= "Cysteine rich repeat"
FT	Region	691..746	/label= CR_4
FT	Region	765..820	/note= "Cysteine rich repeat"
FT	Region	765..820	/label= CR_5
FT	Region	831..885	/note= "Cysteine rich repeat"
FT	Region	831..885	/label= CR_6
FT	Region	831..885	/note= "Cysteine rich repeat"
PN	WO200138519-A1.		
PD	31-MAY-2001.		
XX	24-NOV-2000; 2000WO-AU01435.		
PR	26-NOV-1999; 99AU-0004348.		
PA	(UYQU) UNIV QUEENSLAND.		
PI	Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;		
DR	WPI; 2001-343951/36.		
DR	N-PSDB; AAS11603.		
PT	Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,		
PT	useful for preventing, diagnosing and treating e.g. eye disease,		
PT	especially cataract formation -		
XX	Claim 11; Fig 1; 169pp; English.		
XX	The invention relates to nucleic acids from human chromosome 2p21-16.3		
CC	and the encoded peptide (and mouse and chicken orthologues) that		
CC	comprises a PGECCPLP group, an insulin-like growth factor binding protein		
CC	(IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group		
CC	and a transmembrane domain. The protein, e.g. CRIM1, interacts with		
CC	peptides of the transforming growth factor superfamily. A composition		
CC	comprising an expression construct comprising the nucleic acids of the		
CC	invention or a mimetic which antagonises or mimics an activity of a CRIM1		
CC	polypeptide may be used in a method for modulating the biological		
CC	activity of a polypeptide of the bone morphogenic protein (BMP) family.		
CC	In this way they may be used to prevent or treat an eye disease,		
CC	especially cataract formation. They may also be used to treat		
CC	neurodegenerative diseases, renal and kidney disease, bone and tooth		
CC	abnormalities, wounds and skin damage, e.g. by use of the nucleic acid in		
CC	gene therapy by using antibodies directed against CRIM1 polypeptides.		
CC	The present sequence represents chicken CRIM1 (AKA S52).		
XX			
XX	Sequence 1048 AA:		
SO			
Query Match	16.9%; Score 163; DB 22; Length 1048;		
Best Local Similarity	45.1%; Pred. No. 8.9e-07;		
Matches 23; Conservative 11; Mismatches 17; Indels 0; Gaps 0			
OY	117 GESHANNTKKKKDACTICECKDQGVTCVEACPRATCAVPVNIIPACPCVC 167		
Db	835 GKTVADEERWDIDSCTHCYCLOGGTLCTVSCPPPLPCAEPINVEGSCPCMC 885		
RESULT 13			
ID	ABU70387 standard; Protein: 293 AA.		

AC AB070387;
XX
XX 10-JUN-2003 (first entry)
DT XX
DE Human adipocyte Selected Interacting domain, SID, #18.
XX
KW Human; prey; adipocyte; SID; selected interacting domain;
XX anorectic; antidiabetic; protein-protein interaction; diabetes;
XX yeast 2-hybrid assay; metabolic disorder; obesity.
OS Homo sapiens.
XX
XX W0200286122-A2.
PN
XX
PD 31-OCT-2002.
PE
XX 14-MAR-2002; 2002W0-EPO3768.
PR 14-MAR-2001; 2001US-275734P.
XX
XX (HYBR-) HYBRIGENICS.
PA
PI Legrain P, Davlet L;
DR WPI: 2003-103412/09.
DR N-PsDB: ACA56931.
XX
XX New complex between two interacting proteins in adipocyte cells, useful
PT for identifying selected interacting domains that modulate protein
PT interactions, or for preventing or treating metabolic disorders such as
PT obesity or diabetes -
PS
PS Claim 6; Page 114; 382pp; English.

The invention relates to a complex between two interacting proteins in
CC adipocyte cells, given in the specification. The proteins are identified
CC by selecting a bait protein from a known adipocyte marker and then
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
CC members of an adipocyte cDNA library. The proteins are designated SID
CC (RTM) (selected interacting domains) proteins. Also included are a
CC polynucleotide encoding a polypeptide in the adipocyte cells, a
CC recombinant host cell expressing at least one of the interacting
CC polypeptides of the complex, selecting a modulating compound in adipocyte
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
CC sequences given in the specification (including its fragment or variant),
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
CC given in the specification (including its fragment or variant), a vector
CC comprising the SID (RTM) polynucleotide, a recombinant host cell
CC comprising the vector, a protein chip comprising the polypeptides and
CC a record comprising all or part of the data, listed in the specification.
CC The complex, polypeptides, polynucleotides and compounds are
CC useful for preventing or treating metabolic disorders such as obesity
CC or diabetes. The polynucleotides are useful as probes or primers. The
CC complex is particularly useful for identifying selected interacting
CC domains (SID (RTM)) for screening drugs that modulate the protein
CC interaction, thus exhibiting the therapeutic effect. The present
CC sequence represents a SID (prey) protein of the invention.

Sequence 293 AA:

Query Match	16.6%;	Score 160;	DB 24;	Length 293;
Best Local Similarity	43.1%;	Pred. NO. 3.4e-07;		
Matches 22;	Conservative 11;	Mismatches 18;	Indels 0;	Gaps 0;

```

117 GESNANNTKMKKDDACTICECKDGOVTCVEACRPATCAVPNNIPACCPCVC 167
I::: I: I:I I: I: I: I: I: I: I: I: I: I: I: I: I: I: I:
91 GKAAVEDERMDJDSCTHCYCLGGQTLCTSVSCPPICVEPINIVGSSCCPMC 141

```

RESULT 14
ID AAY82775 standard; Protein: 400 AA
AAY82775

XX AAY82775;
 AC 19-JUN-2000 (first entry)
 DT
 XX
 DE Human chordin related protein (Clone dj167_2).
 XX
 XX Chordin related protein; cartilage; bone; connective tissue;
 XX periodontal disease; osteoporosis; burn; incision; ulcer; neuron;
 XX astrocyte; glial cell; transplantation; nerve; epidermis; muscle;
 XX liver; brain; lung; cardiac; pancreas; kidney; growth;
 XX differentiation; TGF-Beta; angiogenesis; chemotaxis;
 XX chemotactic; collagen synthesis; fibrosis; cell adhesion;
 XX cell migration; fertility; reproduction; hematopoiesis;
 XX erythroid cell; tumour; dietary supplement; growth medium.
 XX
 OS Homo sapiens.
 XX
 XX WO200009551-A1.
 XX
 XX 24-FEB-2000.
 XX
 XX 10-AUG-1999; 99WO-US18117.
 XX
 XX 10-AUG-1998; 98US-0095880.
 XX 06-MAY-1999; 99US-0306111.
 XX
 XX (GEMV) GENETICS INST INC.
 XX
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Merberg D;
 XX Treacy M, DiBlasio-Smith E, Widom A;
 XX
 XX WPI: 2000-205978/18.
 XX N-PSDB; AA293171.
 XX
 PT New polynucleotides encoding secreted human proteins, useful for
 PT treating e.g. broken bones, craniofacial defects, periodontal disease,
 PT osteoporosis, burns, incisions or ulcers -
 XX
 PS Claim 10; Page 92-93; 105pp; English.
 XX
 CC The human chordin related protein and polynucleotides encoding them
 CC are predicted to have biological activities which would make them
 CC suitable for treating, preventing or ameliorating medical conditions
 CC which involve defects in cartilage, bone or connective tissue
 CC formation and damage to cartilage, bone or connective tissue, e.g.
 CC broken bones, congenital, trauma-induced, or
 CC oncologic-resection-induced craniofacial defects, periodontal
 CC disease, defects in the periodontal ligament or attachment apparatus,
 CC damage to the periodontal ligament or attachment apparatus,
 CC osteoporosis, burns, incisions or ulcers. The proteins may also
 CC affect neuronal, astrocytic, and glial cell survival and therefore be
 CC useful in transplantation and treatment of conditions exhibiting a
 CC decrease in neuronal survival and repair. The proteins may also be
 CC useful for the treatment of conditions related to other types of
 CC tissue, such as nerve, epidermis, muscle, and other organs such as
 CC liver, brain, lung, cardiac, pancreas, and kidney tissue. The
 CC proteins may further be useful for the treatment of relatively
 CC undifferentiated cell populations, such as embryonic cells, or stem
 CC cells, to enhance growth and/or differentiation of the cells.
 CC The proteins may also have other useful properties characteristic of
 CC the TGF-beta superfamily of proteins. Such properties include
 CC angiogenic, chemotactic, and/or chemottractant properties, and
 CC effects on cells including induction or inhibition of collagen
 CC synthesis, fibrosis, differentiation responses, cell proliferative
 CC responses, and responses involving cell adhesion, migration, and
 CC extracellular matrices. These properties make the proteins potential
 CC agents for wound healing, reduction of fibrosis, and reduction of
 CC scar tissue formation. Chordin-related proteins may also be useful
 CC for advancement of the onset of fertility in sexually immature
 CC mammals, so as to increase the lifetime reproductive performance of
 CC domestic animals such as cows, sheep and pigs. Chordin-related
 CC proteins may also be useful in modulating hematopoiesis by inducing

CC the differentiation of erythroid cells, for suppressing the
 CC development of gonadal tumors, or for augmenting the activity of
 CC BMPs. The proteins may also have value as a dietary supplement, or
 CC as a component of cell culture media.
 XX
 XX SQ Sequence 400 AA;
 XX
 XX Query Match 16.6%; Score 160; DB 21; Length 400;
 XX Best Local Similarity 43.1%; Pred. No. 5, 1e-07;
 XX Matches 22; Conservative 11; Mismatches 18; Indels 0; Gaps 0;
 XX
 QY 117 GESHANNTKKKDACTICEKDGQVTCFVEACPPATCAVPVNPACCPVC 167
 DB 187 GRAYADEERWDDSCHTCYCLOGQTLCTSVCPPIVCEPIVWESCCPMC 237
 XX
 XX
 XX RESULT 15
 XX AAY53033
 XX ID AAY53033 standard; Protein; 400 AA.
 XX
 XX AC AAY53033;
 XX
 XX DT 29-FEB-2000 (first entry)
 XX
 XX DE Human secreted protein clone dj167_2 protein sequence SEQ ID NO:72.
 XX
 XX Human: secreted protein; nutritional; cytokine; cell proliferation;
 XX differentiation; immune stimulating; vaccine; suppression;
 XX haematopoiesis regulation; tissue growth; activin; inhibin;
 XX chemotactic; chemokinetic; haemostatic; thrombolytic; receptor;
 XX ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
 XX tumour inhibition; gene therapy.
 XX
 XX
 OS Homo sapiens.
 XX
 XX WO9957132-A1.
 XX
 XX 11-NOV-1999.
 XX
 XX 07-MAY-1999; 99WO-US09970.
 XX
 XX 07-MAY-1998; 98US-0084564.
 XX 02-JUN-1998; 98US-0087645.
 XX 22-JUL-1998; 98US-0093712.
 XX 31-JUL-1998; 98US-0094935.
 XX 10-AUG-1998; 98US-0095880.
 XX 11-AUG-1998; 98US-0096068.
 XX 06-MAY-1999; 99US-0096068.
 XX
 XX (GEMV) GENETICS INST INC.
 XX
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Bowman MR;
 XX DiBlasio-Smith E, Widom A;
 XX
 XX WPI: 2000-052937/04.
 XX N-PSDB; AA233351.
 XX
 PT New polynucleotides encoding secreted human proteins, derived from
 PT adult placenta, adult retina, fetal brain, fetal -
 XX
 PS Claim 81; Page 423-424; 492pp; English.
 XX
 CC The present invention describes new human secreted proteins which were
 CC isolated from adult placenta, adult retina, foetal brain, foetal kidney,
 CC adult blood, adult brain, adult thyroid, adult bladder, adult neural
 CC tissue, adult testes, and adult lymph node cDNA libraries. The human
 CC secreted proteins, and the polynucleotides encoding them, are predicted
 CC to have biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals. Suggested activities include nutritional activity, cytokine
 CC and cell proliferation/differentiation activity, immune stimulating
 CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating

CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC cadherin/tumour invasion suppressor activity, and tumour inhibition
 CC activity. The polynucleotides are also stated to be useful for gene
 CC therapy. AA23316 to AA23373 encode human secreted proteins, and
 CC AA52998 to AA53060 represent human secreted proteins, given in the
 CC present invention.

XX
 SQ Sequence 400 AA;

Query Match 16.6%; Score 160; DB 21; Length 400;
 Best Local Similarity 43.1%; Pred. No. 5.1e-07;
 Matches 22; Conservative 11; Mismatches 18; Indels 0; Gaps 0;

QY 117 GESHANNTKWKDACTICECKDQVTCFVEACPPATCAVPVNIPIGACCPYC 167
 Db 187 GRAYADEERWDLDSCTHCYCLQGOTLCSTVSCPPPCPEPINVEGSCCPMC 237

Search completed: July 24, 2003, 12:51:37
 Job time : 24.6234 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:46:24 ; Search time 39.5854 Seconds
(without alignments)
2086.045 Million cell updates/sec

Title: US-09-884-319A-4

Perfect score: 1645

Sequence: 1 KKGKTEODGYOKPTNKHFT.....DTLEQKYVTLDSASFLLCSC 320

Scoring table: BIOSUM62

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*
15: SP-virus:*
16: SP-bacteriap:*
17: SP-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1343.5	81.7	950	4	Q8NHV8
2	1261.5	76.7	507	11	Q9CT39
3	1261.5	76.7	949	11	Q8R266
4	1158.5	70.4	958	13	Q8APX2
5	967.5	58.8	857	11	Q8C043
6	956.5	58.1	867	13	Q90X49
7	798	48.5	385	11	Q9JRW4
8	702.5	42.7	209	11	Q9D624
9	253	15.4	85	4	Q8LV76
10	159.5	9.7	594	4	Q8NBV1
11	155.5	9.5	601	13	Q8APX3
12	147.5	9.0	410	4	Q9H4R2
13	147.5	9.0	465	4	Q60687
14	147.5	9.0	465	4	Q8W85
15	141.5	8.6	467	11	Q8K4W6
16	141.5	8.6	467	11	Q8R054

17	141.5	8.6	468	11	Q8K1F8	Q8K1F8 mus musculu
18	128.5	7.8	380	11	Q9R0M2	Q9R0M2 mus musculu
19	128.5	7.8	464	11	Q9R0M3	Q9R0M3 mus musculu
20	106	6.4	375	17	Q96Y62	Q96Y62 sulfolobus
21	105.5	6.4	735	4	Q96Y63	Q96Y63 homo sapien
22	105.5	6.4	735	4	Q96Y63	Q96Y63 homo sapien
23	103	6.3	435	13	Q07427	Q07427 carassius a
24	102	6.2	839	16	Q97KW9	Q97KW9 clostridium
25	101.5	6.2	353	16	Q9K1E3	Q9K1E3 vibrio chol
26	101	6.1	499	10	Q8IVC0	Q8IVC0 homo sapien
27	101	6.1	1487	10	Q9FH23	Q9FH23 arabidopsis
28	100	6.1	489	16	Q9CHN2	Q9CHN2 lactococcus
29	100	6.1	1012	16	Q8CRN4	Q8CRN4 staphylococ
30	99.5	6.0	545	16	Q8D3F5	Q8D3F5 wigleswort
31	99.5	6.0	1134	17	Q9UY58	Q9UY58 pyrococcus
32	97.5	5.9	524	16	Q8EX35	Q8EX35 mycoplasma
33	97.5	5.9	790	5	Q9G5F3	Q9G5F3 podocoryne
34	97.5	5.9	798	2	Q9EV58	Q9EV58 spiroplasma
35	97	5.9	218	17	Q9HP21	Q9HP21 halobacteri
36	97	5.9	388	2	Q8GPH9	Q8GPH9 staphylococ
37	97	5.9	728	17	Q59093	Q59093 pyrococcus
38	97	5.9	1190	5	Q8IU85	Q8IU85 plasmodium
39	96	5.8	397	10	Q9AVY3	Q9AVY3 guillardia
40	96	5.8	434	16	Q9CEC0	Q9CEC0 lactococcus
41	96	5.8	480	5	Q03633	Q03633 plasmodium
42	96	5.8	1081	16	Q8XIU2	Q8XIU2 clostridium
43	95.5	5.8	476	10	Q9ZUK2	Q9ZUK2 arabidopsis
44	95.5	5.8	1170	10	Q8LNR4	Q8LNR4 oryza sativ
45	95.5	5.8	1383	5	Q8I659	Q8I659 plasmodium

ALIGNMENTS

RESULT 1	
Q8NHV8	PRELIMINARY; PRT; 950 AA.
AC Q8NHV8:	
DT 01-OCT-2002 (TREMBLrel. 22, Created)	
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)	
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)	
DE URL:	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA Guo J.H., Yu L.;	
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.	
DR EMBL: AF506819; AAM31633.1; -	
DR SEQUENCE 950 AA; 108186 MW; 67A2C7FC11BFD2A0 CRC64;	

Query Match 81.7%; Score 1343.5; DB 4; Length 950;
Best Local Similarity 81.4%; Pred. No. 1.6e-99;
Matches 267; Conservative 0; Mismatches 0; Indels 61; Gaps 1;

QY	1 KKGKTEODGYOKPTNKHFTQSPKKSVDLLGSGFKRRLLITAPKKNMYVOQREY 60
DB	566 KKGKTEODGYOKPTNKHFTQSPKKSVDLLGSGFKRRLLITAPKKNMYVOQREY 645
QY	61 LESFCKMATRKISVTITFGPVNNSYMKIDHFDLNEXKPMRVYDDELDVDRISLRLKEY 120
DB	646 LESFCKMATRKISVTITFGPVNNSYMKIDHFDLNEXKPMRVYDDELDVDRISLRLKEY 705
QY	121 GMTYNDFFVNLTDVLDLVKQYQYEVPIYTKSVFDLIDFQSRITKDEKQKGGIVCK --- 176
DB	706 GMTYNDFFVNLTDVLDLVKQYQYEVPIYTKSVFDLIDFQSRITKDEKQKGGIVCKEDKK 765
QY	177 ----- -DEV 179
DB	766 OSLENFLSRFRRLRLVISAPEDEDMAYSQQLSALSGOACNFGRLHITIKLLGAGREV 825

OY 180 GGVLELFPINGSSVVEREDVPAHLVKDIRNFQVSPPEFSSMLLVKDGKGVKSWPSPMWS 239
 DB 826 GGVLELFPINGSSVVEREDVPAHLVKDIRNFQVSPPEFSSMLLVKDGKGVKSWPSPMWS 885
 OY 240 MVIYVDLIDSMQLRROEMAIQOOSLGMR 267
 DB 886 MVIYVDLIDSMQLRROEMAIQOOSLGMR 913

RESULT 2

OCT39 PRELIMINARY: PRT: 507 AA.
 ID 09CT39
 AC 09CT39
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE 2610001E17RIK protein (Fragment).
 GN 2610001E17RIK
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shitagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Kawai T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakata I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Watanabe-Borja A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection."
 RT Nature 409:685-690(2001).
 RL EMBL: AK011256; BAB27498.1; -
 DR MGD: MGI:1915146; 2610001E17RIK.
 FT NON_TER
 SQ SEQUENCE 507 AA: 58718 MW; EDCFODB003ECF192 CRC64;

Query Match 76.7%; Score 1261.5; DB 11; Length 507;
 Best Local Similarity 76.2%; Pred. No. 2.9e-93;
 Matches 250; Conservative 6; Mismatches 11; Indels 61; Gaps 1;

OY 1 KKGKTEODGYOKPTNKHFTOSPKKSVADLLGSFEGKRRLITAPKAENMYVOQRDEY 60
 DB 143 KKGKTEODDQKPTAKHLAPSPKSVADLLGSFEGKRRLITTPKAENMYVOQRDEY 202
 OY 61 LESFCKMATRKISVTTITGPNVNSMTKIDHQLDNEKPMRVYDDDLVDQRLISELRKEX 120
 DB 203 LESFCKMATRKISVTTITGPNVNSMTKIDHQLDNEKPMRVYDDDLVDQRLISELRKEX 262
 OY 121 GMTYNDFFMVLTVDVLRKQYEVPIITMKSVDLIDTFQSRITKIDMEKKEGIVCK---- 176
 DB 263 GMTYNDFFMVLTVDVLRKQYEVPIITMKSVDLIDTFQSRITKIDMEKKEGIVCKEDKR 322
 OY 177 -----EEV 179
 DB 323 QSLNLFSLRFRMRRLIVISAPNDEDMVYSQOLALNGQACNFGRLHTITLLKLGVEEV 382
 OY 180 GGVLELFPINGSSVVEREDVPAHLVKDIRNFQVSPPEFSSMLLVKDGKGVKSWPSPMWS 239
 DB 886 MVIYVDLIDSMQLRROEMAIQOOSLGMR 913

DB 383 GGVLELFPINGSSVVEREDVPAHLVKDIRNFQVSPPEFSSMLLVKDGKGVKSWPSPMWS 442
 OY 240 MVIYVDLIDSMQLRROEMAIQOOSLGMR 267
 DB 443 MVIYVDLIDSMQLRROEMAIQOOSLGMR 470

RESULT 3

OBR2G6 PRELIMINARY: PRT: 949 AA.
 ID OBR2G6
 AC OBR2G6
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE URB precursor.
 GN 2610001E17RIK OR URB.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=White adipose tissue;
 RX MEDLINE=21670972; PubMed=11812002;
 RA Aoki K., Sun Y., Aoki S., Wada K., Wada E.;
 RT "Cloning, expression, and mapping of a gene that is upregulated in
 RT adipose tissue of mice deficient in bombesin receptor subtype-3."
 RL Biochem. Biophys. Res. Commun. 290:1282-1288(2002).
 DR EMBL: AB075019; BAB5613.1; -
 DR MGD: MGI:1915146; 2610001E17RIK.
 KW Signal.
 FT SIGNAL
 SQ SEQUENCE 949 AA: 107640 MW; 62693C715C36F6AB CRC64;

Query Match 76.7%; Score 1261.5; DB 11; Length 949;
 Best Local Similarity 76.2%; Pred. No. 6.5e-93;
 Matches 250; Conservative 6; Mismatches 11; Indels 61; Gaps 1;

OY 1 KKGKTEODGYOKPTNKHFTOSPKKSVADLLGSFEGKRRLITAPKAENMYVOQRDEY 60
 DB 585 KKGKTEODDQKPTAKHLAPSPKSVADLLGSFEGKRRLITTPKAENMYVOQRDEY 644
 OY 61 LESFCKMATRKISVTTITGPNVNSMTKIDHQLDNEKPMRVYDDDLVDQRLISELRKEX 120
 DB 645 LESFCKMATRKISVTTITGPNVNSMTKIDHQLDNEKPMRVYDDDLVDQRLISELRKEX 704
 OY 121 GMTYNDFFMVLTVDVLRKQYEVPIITMKSVDLIDTFQSRITKIDMEKKEGIVCK---- 176
 DB 705 GMTYNDFFMVLTVDVLRKQYEVPIITMKSVDLIDTFQSRITKIDMEKKEGIVCKEDKR 764
 OY 177 -----EEV 179
 DB 765 QSLNLFSLRFRMRRLIVISAPNDEDMVYSQOLALNGQACNFGRLHTITLLKLGVEEV 824
 OY 180 GGVLELFPINGSSVVEREDVPAHLVKDIRNFQVSPPEFSSMLLVKDGKGVKSWPSPMWS 239
 DB 825 GGVLELFPINGSSVVEREDVPAHLVKDIRNFQVSPPEFSSMLLVKDGKGVKSWPSPMWS 884
 OY 240 MVIYVDLIDSMQLRROEMAIQOOSLGMR 267
 DB 885 MVIYVDLIDSMQLRROEMAIQOOSLGMR 912

RESULT 4

OBRXP2 PRELIMINARY: PRT: 958 AA.
 ID OBRXP2
 AC OBRXP2
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Equartin-L precursor.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OY
I KKGKTEODGYOKPTNKHFTOSEPKSVADLLGSEFGKRLLLTTPAKAENNNYYVOORDEX 60

RESULT 7
Q9JKM4
ID Q9JKM4 PRELIMINARY; PRT; 385 AA.

AC Q9JRW4; 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Steroid sensitive gene-1 protein.
 GN SSG-1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21255622; PubMed=11356689;
 RA Marcantonio D., Chailifour L.E., Alaoui-Jamali M.A., Alpert L.,
 RA Huyhn H.T.;
 RT "Cloning and characterization of a novel gene that is regulated by
 RT estrogen and is associated with mammary gland carcinogenesis.";
 RL Endocrinology 142:2409-2418(2001).
 DR EMBL; AF23677; AAF35351.1;
 SO SEQUENCE 385 AA; 44066 MW; 461746FE7837346 CRC64;

Query Match 48.5%; Score 798; DB 11; Length 385;
 Best Local Similarity 84.0%; Pred. No. 4,1e-56;
 Matches 158; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

OY 1 KKGKTEQDDGYQKPTNKHFTQSPKRSVADLLGSFEGKRRLLTTPAKAENMYVOQRDEY 60
 DB 183 KKAGTEQDDGYQKPTAKHLAPSPKRSVADLLGSFEGKRRLLTTPKAKENMYVOQRDEY 242
 OY 61 LESFCKMATRKISVTTTIGCPYNNSTMKIDHFDQDNKEMRYVDDDLVDQRLISLRKEY 120
 DB 243 LESFCKMATRKISVTTTIGCPYNNSTMKIDHFDQDNKEMRYVDDDLVDQRLISLRKEY 302
 OY 121 GMTYDFEFWNLTDVLRKQYEVPTMKSYEDLDTFQSRKDKMEKQKKEGYCEVNG 180
 DB 303 GMTYDFEFWNLTDVLRKQYEVPTMKSYEDLDTFQSRKDKMEKQKKEGYCEVNG 362
 OY 181 GVLELFPIT 188
 DB 363 AVPGGEFPI 370

RESULT 8
 O9D624 PRELIMINARY; PRT; 209 AA.
 AC Q9D624;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 2610001E17RIK protein.
 GN 2610001E17RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK009795; BAB26508.1;
 DR MGI; MGI:1915146; 2610001E17RIK.
 SO SEQUENCE 209 AA; 24334 MW; 9964F0FB03ABB8C3 CRC64;

Query Match 42.7%; Score 702.5; DB 11; Length 209;
 Best Local Similarity 68.6%; Pred. No. 9e-49;
 Matches 142; Conservative 2; Mismatches 2; Indels 61; Gaps 1;

OY 122 MTYNDFFWNLTDVLRKQYEVPTMKSYEDLDTFQSRKDKMEKQKKEGYCK----- 176
 DB 1 MTYNDFFWNLTDVLRKQYEVPTMKSYEDLDTFQSRKDKMEKQKKEGYCKEKKRQ 60
 OY 177 -----EYNG 180
 DB 61 SLENFLSFRWRRLVISAPEDEDMAYSQLSALNGQACNFGIRHTITIKLGVGEVNG 120
 OY 181 GVLELFPITNGSSYVEREDVPAHLVKDIRNFQVSPETFSMLLVGKDGKSWPSPMSM 240
 DB 121 GVLELFPITNGSSYVEREDVPAHLVKDIRNFQVSPETFSMLLVGKDGKSWPSPMSM 180
 OY 241 VIVVDLIDSMQLRQEMAIQOSLGMR 267
 DB 181 VIVVDLIDSMQLRQEMAIQOSLGMR 207

RESULT 9
 O81VT6 PRELIMINARY; PRT; 85 AA.
 AC O81VT6;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Similar to RIKEN CDNA 2610001E17 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RL Strausberg R.;
 DR EMBL; BC042105; AAH42105.1;
 SO SEQUENCE 85 AA; 10174 MW; F4381D3541675A0 CRC64;

Query Match 15.4%; Score 253; DB 4; Length 85;
 Best Local Similarity 100.0%; Pred. No. 4.2e-13;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 220 MLVKGKDGKSWPSPMSMVIYVDLIDSMQLRQEMAIQOSLGMR 267
 DB 1 MLVKGKDGKSWPSPMSMVIYVDLIDSMQLRQEMAIQOSLGMR 48

RESULT 10
 O8NBV1 PRELIMINARY; PRT; 594 AA.
 AC O8NBV1;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein FLJ90729 (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RC Tissue-Placenta;
 RA Isocai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
 RA Masuno Y., Ota T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Nihomiya K.,
 RT "NEO human cDNA sequencing project";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK075210; BAC11475.1; -
 KM Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 594 AA; 66667 MW; BF55F9E9C6783F6B CRC64;

Query Match 9.7%; Score 159.5; DB 4; Length 594;
 Best Local Similarity 29.3%; Pred. No. 0.00018;
 Matches 53; Conservative 35; Mismatches 64; Indels 29; Gaps 8;

QY 26 SVADLLGSEGRRLITLTPAKENNNY-----VOQREYLESEFCMKATRKISVTITGP 80
 DB 148 SSPNLTAFACKNRWVWISAPHASGYYRLMMSLKLDVY---CELAERHIQIVLPHQ 203
 QY 81 VVNSMTKIDHFDLND---EKPRVYDDEDVLDQRLSELKREYGMTYDFEAVLTVDVLR 137
 DB 204 AGEEGKAVRRTTSEGLLEOPL---DPSLI-PKIMSFLKLEKG---KFGVYLKKTLO 254
 QY 138 VKOYEVPTMKSVFDLIDFOSRIKMEKKEGIVCKEYGVLELFPINGSSVVERE 197
 DB 255 VEERTPYRLEAMTEVID--QGPIRIETIRKQGFVOKCAGV-----EGOVVAEGN 306
 QY 198 D 198
 DB 307 D 307

RESULT 11

Q8AXP3 PRELIMINARY; PRT; 601 AA.
 AC Q8AXP3;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Equarlin-5 precursor.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hong M., Ohta K., Kuriyama S., Tanihara H., Yasuda K., Tanaka H.;
 RT "Equarlin: novel soluble molecules expressed in the equator region of
 chick embryonic lens";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB086823; BAC54278.1; -
 KM Signal.
 FT SIGNAL
 SQ SEQUENCE 601 AA; 67676 MW; 7AF1A2A518A4152A CRC64;

Query Match 9.5%; Score 155.5; DB 13; Length 601;
 Best Local Similarity 21.7%; Pred. No. 0.00037;
 Matches 74; Conservative 53; Mismatches 119; Indels 95; Gaps 13;

QY 2 KGGKTEQDGYQPTKHTQSPK-KSVADLLGSEGRRLITLTPAKENNNY-----VO 55
 DB 112 KSEMTKDEGISTASOSRAVRPSSGSSPVVLASFGRKRWVWISAPHASGYYRLMMSL 171
 QY 56 QDEYLESEFCMKATRKISVTITGPVNNSTMKIDHFDLNDKPRVYDDEDVLDVQ----- 110
 DB 172 KNDVY---CELAERHIQIVLFHEGE-----EGGVVRIITNGKTLLEPDP 217
 QY 111 ---RLSEKREYGMTYDFEAVLTVDVLRKQYEVPTMKSVFDLIDFOSRIKMEK 167

DB 218 LIPKLMSEFLKLEK----KFGVLLKKTQVEERYPYRLEAMEVID--QNPLRIK 271
 QY 168 OKKEIV--CKEEVGVLELPPINGSSVVEEDPAHVKQIRNFQVSPFYSLVYK 225
 DB 272 MKQKGFQTK-----AAGEGVQVED----- 293
 QY 226 DGNVSWYPSPMWMSVIYVYDIIDLSQRLROEMAIQOISGMCKMSQAMVTI----- 278
 DB 294 DNGSGTOSITGGHGVQY-----SAGGRKEPRSSNQPTRTKYRKMTTVAATPLPTV 348
 QY 279 -----VTKDTRMVTMTYIMRYTMDT---LTEQKRYT 310
 DB 349 RTTLPTTTATRTATRTATRTATRTATRTATRTATRTATRTATRTATRTATRTAT 389

RESULT 12

Q9H4R2 PRELIMINARY; PRT; 410 AA.
 AC Q9H4R2;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE BA524DI6A.1 (Sushi-repeat-containing protein) (Fragment).
 GN SRPX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilson S.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL391688; CAC16060.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR003410; HyalIn.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF02494; HYR; 1.
 DR Pfam: PF00084; sushi; 3.
 DR SMART: SM00032; CCP; 3.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 FT NON_TER
 SQ SEQUENCE 410 AA; 46781 MW; D32228E1AA1AE370 CRC64;

Query Match 9.0%; Score 147.5; DB 4; Length 410;
 Best Local Similarity 26.5%; Pred. No. 0.001;
 Matches 40; Conservative 29; Mismatches 67; Indels 15; Gaps 1;

QY 26 SVADLLGSEGRRLITLTPAKENNNYVOQREYLESEFCMKATRKISVTITGPVNNST 85
 DB 275 SAAGLLDQFEYKQRLITISAPDSNRYKKQISMLOSTGIDLHVTITELVQGPQEV 334
 QY 86 MKIDHFDLNDKPRVYDDEDVLDQRLSELKREYGMTYDFEAVLTVDVLRKQYEV 145
 DB 335 GRIEQQ-----SANTIEELRQFORLRSFYNNVILDKGIDRQYMEP 379
 QY 146 ITKSVFLPIDFOSRIKMEKKEGIVCK 176
 DB 380 VTPEITFTFIDYLLSNOELVQRQRDICE 410

RESULT 13

Q60687 PRELIMINARY; PRT; 465 AA.
 AC Q60687;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Sushi-repeat protein (Sushi-repeat containing protein).
 GN SRPV.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Kurosawa H., Inukai T., Inaba T., Goi K., Chang K.-S., Shinjo T.,
 RA Rakestraw K.M., Naeye C.W., Look T.A.;
 RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Huang C.-H., Chen H., Peng J., Chen Y.;
 RT "Cloning and characterization of the sushi-repeat containing protein
 RT (SRP) as a novel interaction partner of Rh type C glycoprotein
 RT (RhCG).";
 RL Submitted (Jun-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF060567; AAC15765.1; -;
 DR EMBL: AF393649; AAM73693.1; -;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR003410; HyalIn.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF02494; HVR; 1.3.
 DR Pfam: PF00084; sushi; 3.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 SQ SEQUENCE 465 AA; 52971 MW; 4D752B187FF3EFB8 CRC64;

Query Match 9.0%; Score 147.5; DB 4; Length 465;
 Best Local Similarity 26.5%; Pred. No. 0.0012;
 Matches 40; Conservative 29; Mismatches 67; Indels 15; Gaps 1;

QY 26 SVADLLGSEFGKRRLLITAPKAENNMVVOQRDEYLESEFCMKATRKISVITTFGPNVNST 85
 DB 330 SAAGLLDGFYEKQRLIISAPDPSNRYKMQISMLOSTGGLDLRHVITIELVGPQPEV 389
 QY 86 MKIDHFQLDNEKPMRVNDEDLVDQRLISELRKEYGMYNDFFWLTVDLRVQYVEVP 145
 DB 390 GRIRFQQL-----SANIIEELRQFQRLRSYFNMYLIDKQIDRDRYMEP 434
 QY 146 ITMKSVFPLIDTFQSRIDMEKQKEGIVCK 176
 DB 435 VTPEEIFTFIDYLLSNQELTORRORDICE 465

RESULT 14
 O8W85 PRELIMINARY; PRT; 465 AA.
 AC O8W85;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE Sushi-repeat protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Plimates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Strausberg R.;
 RL Submitted (Jan-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC020733; AAH20733.1; -;
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR003410; HyalIn.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF02494; HVR; 1.3.
 DR Pfam: PF00084; sushi; 3.
 DR SMART: SM00032; CCP; 3.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 SQ SEQUENCE 465 AA; 52957 MW; 3D7229487DA1B8BD CRC64;

Query Match 9.0%; Score 147.5; DB 4; Length 465;
 Best Local Similarity 26.5%; Pred. No. 0.0012;
 Matches 40; Conservative 29; Mismatches 67; Indels 15; Gaps 1;

QY 26 SVADLLGSEFGKRRLLITAPKAENNMVVOQRDEYLESEFCMKATRKISVITTFGPNVNST 85
 DB 330 SAAGLLDGFYEKQRLIISAPDPSNRYKMQISMLOSTGGLDLRHVITIELVGPQPEV 389

QY 86 MKIDHFQLDNEKPMRVNDEDLVDQRLISELRKEYGMYNDFFWLTVDLRVQYVEVP 145
 DB 390 GRIRFQQL-----SANIIEELRQFQRLRSYFNMYLIDKQIDRDRYMEP 434
 QY 146 ITMKSVFPLIDTFQSRIDMEKQKEGIVCK 176
 DB 435 VTPEEIFTFIDYLLSNQELTORRORDICE 465

RESULT 15

O8K46 PRELIMINARY; PRT; 410 AA.
 AC O8K46;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DE Sushi-repeat containing protein (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Huang C.-H., Chen H., Peng J., Chen Y.;
 RT "Cloning and characterization of the sushi-repeat containing protein
 RT (SRP) as a novel interaction partner of Rh type C glycoprotein
 RT (RhCG).";
 RL Submitted (Jun-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF393647; AAM73690.1; -;
 DR EMBL: AF393641; AAM73690.1; JOINED.
 DR EMBL: AF393642; AAM73690.1; JOINED.
 DR EMBL: AF393643; AAM73690.1; JOINED.
 DR EMBL: AF393644; AAM73690.1; JOINED.
 DR EMBL: AF393645; AAM73690.1; JOINED.
 DR EMBL: AF393646; AAM73690.1; JOINED.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR003410; HyalIn.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF02494; HVR; 1.3.
 DR Pfam: PF00084; sushi; 3.
 DR SMART: SM00032; CCP; 3.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 FT NON_TER 1
 SQ SEQUENCE 410 AA; 46804 MW; ICEE27C10132D04B CRC64;

Query Match 8.6%; Score 141.5; DB 11; Length 410;
 Best Local Similarity 25.8%; Pred. No. 0.003;
 Matches 39; Conservative 30; Mismatches 67; Indels 15; Gaps 1;

QY 26 SVADLLGSEFGKRRLLITAPKAENNMVVOQRDEYLESEFCMKATRKISVITTFGPNVNST 85
 DB 275 SAAGLLDGFYEKQRLIISAPDPSNRYKMQISMLOSTGGLDLRHVITIELVGPQPEV 334
 QY 86 MKIDHFQLDNEKPMRVNDEDLVDQRLISELRKEYGMYNDFFWLTVDLRVQYVEVP 145
 DB 335 GRIRFQQL-----SAGIIEELRQFQRLRSYFNMYLIDKQIDRDRYMEP 379
 QY 146 ITMKSVFPLIDTFQSRIDMEKQKEGIVCK 176
 DB 380 VTPEEIFTFIDYLLSNQELTARVQRDICE 410

Search completed: July 24, 2003, 12:56:51
 Job time : 42.5854 secs

FT	DISULFID	262	286	BY SIMILARITY.		
FT	DISULFID	275	290	BY SIMILARITY.		
SO	SEQUENCE	464 AA:	51559 MM:	E27C5B163090A97E CR664:		
Query Match						
Best Local Similarity		7.9%;	Score 130.5;	DB 1; Length 464;		
Matches		42;	Conservative 25;	Mismatches 69; Indels 15; Gaps 1		
QY	25	KSVA	LDLGSFEGKRLRLITAPKPA	NNNNYYOORDEYLSFCKMARTKISVINIFGGVNN	84	
		:::	:::			
Db	327	RTAA	ALLDQFYEKRRLIVSTPTAN	LLIRLQGLMLOQOGLDLRHITV	ELVGVFPL 386	
QY	85	TMK	IDHFO	LDEKPRVYDDELDVDRLLISEL	KEYGATYNDFFVYLDVDVLRVQYEV 144	
Db	387	IGRI	-----	RAKIM	PAPALALQRLRLIRIPLYSF	SVLVDKHMGEREYVS 431
QY	145	PIT	KSVFEDLIDT	QFSRIKDEMKQKKEGIVC	175	
		::: :::	:::			
Db	432	LVT	PALFNLIDT	EPRLKEEMILQEKQSC	462	

RESULT 2			
SRPX_HUMAN			
ID	SRPX_HUMAN	STANDARD:	PRF: 464 AA.
AC	P78339	099652; Q99913;	
DT	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	15-SEP-2003	(Rel. 42, Last annotation update)	
DE	Sushi repeat-containing protein SRPX precursor.		
GN	SRPX OR ETX1.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96177670; PubMed=8634708;		
RA	Meinl A., Carvalhal M.R.S., Herrmann K., Lorenz B., Achatz H.,		
RA	Lorenz B., Apfelfeldt-Sylla E., Wittwer B., Ross M., Meitinger T.,		
RT	A gene (SRPX) encoding a sushi-repeat-containing protein is deleted		
RT	in patients with x-linked retinitis pigmentosa.";		
RL	Hum. Mol. Genet. 4:2339-2346(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A., AND VARIANTS LEU-20 DEL; SER-225 AND PHE-413.		
RX	MEDLINE=96177671; PubMed=8634709;		
RA	Dry R.L., Aldred M.A., Edgar A.J., Brown J., Manson F.D., Ho M.-F.,		
RA	Prosser J., Hadwick L.J., Lennon A.A., Thomson K., van Keuren M.,		
RA	Kurnit D.M., Bird A.C., Jay M., Monaco A.P., Wright A.F.;		
RT	Identification of a novel gene, ETX1 from Xp21.1, a candidate gene		
RT	for x-linked retinitis pigmentosa (RP3).";		
RL	Hum. Mol. Genet. 4:2347-2353(1995).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=97306278; PubMed=9162095;		
RA	Nangaku M., Shankland S.J., Kurokawa K., Bomsztyk K., Johnson R.J.,		
RA	Conser W.G.;		
RT	"Cloning of a new human gene with short consensus repeats using the		
RT	EST database.";		
RL	Immunogenetics 46:99-103(1997).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Placenta;		
RX	MEDLINE=22388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.R., Rubin G.M., Hong L.,		
RA	Simpkins M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.		
RA	Brownstein M.J., Utshin T.B., Toshimaki S., Carninci P., Prange C.,		
RA	Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Wulfsberg S.J.,		
RA	Bosch S.A., McMan P.J., McKernan K.J., Malek J.A., Guarnatone P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L., Hulik S.W.		

[illegible]

```

FT CONFLICT 170 178 RPASCDME -> ASILGCG (IN REF. 3).
FT CONFLICT 202 MISSING (IN REF. 3).
SQ SEQUENCE 464 AA: 51571 MW: 9CDS5E57BC284680 CRC64:

Query Match
Best Local Similarity 26.5%; Score 121.5; DB 1; Length 464;
Matches 40; Conservative 25; Mismatches 71; Indels 15; Gaps 1;

OY 25 KSAVALLGSEFGKRRLLITAPKAENNNVVOORDEYLSFCKMATRKISVITIFGPVNN 84
DB 327 KTAALLDGFYKRRLLITAPKAENNNVVOORDEYLSFCKMATRKISVITIFGPVNN 386
OY 85 TMRIDHFDLNEKPKRVYDDEDLDVQRLISELREKYGMTYNDFFVNLVDVLRVQYEV 144
DB 387 IGR1-----GAKIMPPALALQLRLRLRPLYSFSNVLVQKHGMKDERYVS 431
OY 145 PITMSVFDLITFOSRIKMEKQKKEGIVC 175
DB 432 LVMVALLFNLIDTFPLRREKENVLAQEMSGTC 462

RESULT 3
T2AZ_HUMAN
ID T2AZ_HUMAN STANDARD: PRT; 1182 AA.
AC Q9Y6Q2;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Stomed B-TFIIA-alpha and beta like factor (SALF).
GN SALF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99292779; PubMed=10364255;
RA Upadhyaya A.B., Lee S.H., DeJong J.;
RT "Identification of a general transcription factor TFIIAalpha/beta
homolog selectively expressed in testis.";
RL J. Biol. Chem. 274:18040-18048(1999).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS, HEART, PLACENTA,
KIDNEY, PROSTATE AND UTERUS.
CC -1- SIMILARITY: SOME, IN THE N-TERMINUS TO MEMBERS OF THE ADAPTER
COMPLEXES MEDIUM SUBUNITS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sdb.ch/announce/
or send an email to license@sdb-sib.ch).
CC -----
CC EMBL: AF026169; AAD39617.1; -.
CC DR HSSP: P32773; LYTF.
CC DR MIM: 605357; -.
CC DR GO: GO:0003700; F:transcription factor activity; TAS.
CC DR InterPro: IPR001392; Clathrn_med.
CC DR Pfam: PF00928; Adap_comp_sub; 1.
CC DR Pfam: PF03153; TFIIA; 1.
CC DR Transcription regulation; Nuclear protein.
CC DR DOMAIN 44 150 PRO/SER/THR-RICH.
CC FT DOMAIN 533 536 POLY-VAL.
CC SQ SEQUENCE 1182 AA: 132037 MW: 22AD6366D21C371 CRC64:

Query Match
Best Local Similarity 21.5%; Pred. No. 3.4; Length 1182;
Matches 45; Conservative 43; Mismatches 84; Indels 37; Gaps 8;

```

```

OY 19 FTQSPKSVADLLGSFEGKRR-----LLITAPKAENNNVVOO-----RDEYLESE 64
DB 268 FRSQKSGSMFRLRIPEKKNMSSRWQMGIFLKVLPGLQLQMYQGLEPKKEQLDLPY 327
OY 65 CMAATRKISVITIFGPVNNSTMRIDHFDLNEKPKRVYDDEDLDVQRLISELREKYGMTY 124
DB 328 CRLSEPKVENFSVAKIH--TVKIEHVSYTEKR--KYSKTEVVAHEPDIEQLKLGSTSY 383
OY 125 NOEFVNLVDVLR-----KOYEVPTMSVFDLITFOSRI-----KDMKQKK 170
DB 384 HFDLPLFTVEEELMKLPVSKPKKNYEQETSL-----EIVDNFGVKYKEGFEVASVI 439
OY 171 EGVCKEEYGVLELF-PINGSSVERED 198
DB 440 TQIYCLCFVNGVLEFCFLTINDLELPRKDE 468

RESULT 4
IF2P_PYRHO
ID IF2P_PYRHO STANDARD: PRT; 1044 AA.
AC 058822;
DT 15-DEC-1998 (Rel. 37; Created)
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Probable translation initiation factor IF-2 [Contains: pho Infb
Intein (Pho IF2 Intein)].
GN INFB OR PH1095.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OX Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -1- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
THE BINDING OF THE FORMYL METHIONINE-tRNA TO RIBOSOMES. SEEMS TO
FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
CC -1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -1- MISCELLANEOUS: THE INTEIN INTERRUPTS THE GTP BINDING SITE.
CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
ENDONUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sdb.ch/announce/
or send an email to license@sdb-sib.ch).
CC -----
CC EMBL: AP000005; BAA30194.1; -.
CC DR PIR: H71049; H71049.
CC DR HAMAP: MF_00100; -; 1.
CC DR InterPro: IPR000795; EF_GTPbind.
CC DR InterPro: IPR004161; EFTU_D2.
CC DR InterPro: IPR003586; Hedgehog_hntc.
CC DR InterPro: IPR003587; Hedgehog_hntc.
CC DR InterPro: IPR000178; IF2.
CC DR InterPro: IPR006142; INTEIN.
CC DR InterPro: IPR006141; Intein.

```


RP REVISIONS.
 RA Bull C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
 RA Outenback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
 RA Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.
 RL -1- SIMILARITY: TO M. JANNASCHII M01024.
 CC -1- SIMILARITY: SOME, TO TYPE I RESTRICTION ENZYMES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U67562; AAB99215.1; -
 DR TIGR: MJ1214; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR004473; HsDR.
 DR Pfam: PF04313; HsDR_N; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR TIGRFAMS: TIGR00348; hsdR; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1018 AA: 120992 MW: 97D383541BE128E2 CRC64;
 QY
 Query Match 6.0%; Score 98; DB 1; Length 1018;
 Best Local Similarity 17.7%; Pred. No. 10;
 Matches 66; Conservative 75; Mismatches 123; Indels 108; Gaps 18;
 DB 16 NHEFPOSPK--SVALLSFEKRRLLLTAKANNMVQORDYLESCFMARKT- 72
 333 NAFKDCFKINSVEELKGLVEDIK-----ESENNPNISERGVYLVMMHFKDKLK 384
 QY 73 SVITTEPGVNNSTMKIDHFDLNKPMRVYDDEDLDVQRLISELR----- 117
 385 DFIESG-----SIDKKELLILRDEAHRESGKFAIRKKILKNAIATGTGT 432
 QY 118 -----KEY-----GMTYNDPFMVLTDLVRYKQYVEPTIKSV--FDLIDTFQS 160
 DB 433 PVHKDMSTFEKAYAPQEGEFLDRF-----IESJKEGFTLPLIRVVPEDIKDISEE 488
 QY 161 RIKDM-----EKOKKEGIVKEEYGVLELFT--NGSSVVERED-VPAHLVKDIRNY 210
 DB 489 EIKNIIEKLEFVDEEDADKIVSKKEIAEKIKISDLKSESSIKAEAKYIAEHILEDTENF 548
 QY 211 FQVSPFYFSLVLVGKD-----GNVKSQWYSPWMSWVIY-YDLIDSMQLR 253
 DB 549 -----KFKAMVYAQDRKSCILFKKYLDELAKKIKNY--NEWMTQVITVIHNDVEIE 600
 QY 254 ROEMAIQSLGMCQCKMSQ-----AMVTIYTKDTRMTVTRMTYIMVETMDTL 303
 DB 601 NYKKEIEKKYKKNVDELKKKWTEDFINKENPKILYINKK--LLTGSDAPILKTYIHOF 657
 QY 304 TEOKYVTLDSAS 315
 DB 658 LKD-YILLQASA 668
 RESULT 7
 YK9_YEAST
 ID YK9_YEAST STANDARD; PRT; 1769 AA.
 AC P42945;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 200.0 kDa protein in G2F3-IME2 intergenic region.
 GN YJL109C OR J0808.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 RX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / FY1679;
 RX MEDLINE=96090136; PubMed=7483851;
 RA Rasmussen S.W.;
 RT "A 37.5 kb region of yeast chromosome X includes the SWE1, MER2, GSH1
 RT and CSD3 genes, a TCP-1-related gene, an open reading frame similar
 RT to the DAL80 gene, and a tRNA(Arg).";
 RL Yeast 11:873-883(1995).
 CC -1- SIMILARITY: BELONGS TO THE BAP28 FAMILY.
 CC -1- SIMILARITY: Contains 1 HEAT repeat.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X85021; CA859385.1; -
 DR EMBL: Z49384; CAA89404.1; -
 DR PIR: S53378; S53378.
 DR SGD: S0003645; YJL109C.
 DR GO: GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.
 DR GO: GO:0030515; F:snRNA binding activity; IPI.
 DR GO: GO:0030490; P:processing of 20S pre-rRNA; IMP.
 DR InterPro: IPR000357; HEAT_repeat.
 DR PROSITE: PS50077; HEAT_REPEAT; 1.
 KW Hypothetical protein.
 FT REPEAT 1729 1767
 SQ SEQUENCE 1769 AA: 200080 MW: 064480D1D249B241 CRC64;
 QY
 Query Match 5.9%; Score 96.5; DB 1; Length 1769;
 Best Local Similarity 18.3%; Pred. No. 27;
 Matches 46; Conservative 54; Mismatches 107; Indels 45; Gaps 8;
 DB 81 VNNSTMKIDHFDLNKPMRV---VDDEDLVQRLISELRKYGMTYNDPFVLTDLV 136
 DB 1140 INKTEETDDYDVAVRNRLKAYSVLLDSETSCKIRIRIRREFGTLLEGVLEFINSVEL 1199
 QY 137 RVKQYVEPTIKSVFDLIDTFQSRIKDMKKKEGIVKEEYGVLELFTPIG----- 190
 DB 1200 TFS-----CITSGENEASDS-ETSLSDHTTEKE--ILFKVLGNVLQIIPDEYNAVY 1251
 QY 191 --SSVEREDVPAHLVKDIRNYFQVSPFYFSLVLVGKDGNVKSWYSPWMSWVIYVYDLID 248
 DB 1252 PLISTSTNEDIRHLT-----LVISKFELESEEAIPVNNMKV-LLD 1294
 QY 249 SQOLRROEMAIQSLGRCCKMSQAMVITVTKDIRMTVTRMTYIMVETMDTLTEOKY 308
 DB 1295 RMPLESKSVVISQVI-----LNTMTALVSKYRKKESSILTOALTILATEVSSDME 1346
 QY 309 VTLDSASFSCSC 320
 DB 1347 VKTISLALITNC 1358
 RESULT 8
 Y772_ARCFU
 ID Y772_ARCFU STANDARD; PRT; 349 AA.
 AC Q29486;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AF0772.
 GN AF0772.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;

```

OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirtress E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE001051; AAB90476.1; -
DR PIR; D69346; D69346.
DR TIGR; AF0772; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 349 AA; 40001 MW; A87BCABE90757C4 CRC64;

Query Match 5.8%; Score 96; DB 1; Length 349;
Best Local Similarity 22.6%; Pred. No. 3.8;
Matches 60; Conservative 42; Mismatches 89; Indels 74; Gaps 15;

OY 19 FTQSPKRSYADLLGSEFGKRRLLITAPRAENNMVVOQDEYLESEF--CKMATRKISVIT 76
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 31 FRGNDEVEVEGRSAKSGKGFNLVPL-----FYVDSYSLSSLSRLKMRANVEEAR 84

OY 77 IFG---PVNN-----STKIDHFOLD--NEKPRVYDDELDVQRLISELRKEY 120
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 85 IFKEAIPYINLTLPFRDEKANAVSLNLEPFGIGKPKRIANGSIVSILLITDKNF 144

OY 121 GMTVNDFFVVL-TDVDLRVKQYEVPIPMKSVFDLIDFQSRIDKMEKQKGGIVCKEY 179
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 145 GKIYDFLEPERGTDE-----ETKIIKKEFQTE-----DVAISK----- 180

OY 180 GGVLELFPINGSSVVEREDVPAHLVK--DIRN-----YFOVSPYEFSGMLVGGKDNVK 230
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 181 -GNLSIAKINNMSI--REFLRSHNLRMRGDLNDVNRREGIFGASP--YLLALISKRTNKS 235

OY 231 S-----WYSPMSMVIYVD 245
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 236 TGLGLMDYLNKPYPS-FFTLIDIFYD 259

RESULT 9
TTPL_YEAST STANDARD; PRT; 597 AA.
AC P38069;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TTP1 protein (Mannan synthesis protein MN2).
OS TTP1 OR MN2 OR YBR015C OR YBR0220.
OC Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95084638; PubMed=7992511;
RA Romero P.A., Athanasiadis A., Lussier M., Herscovics A.;
RT "The nucleotide sequence of TTP1, a gene encoding a predicted type II
RT membrane protein.";
RL Yeast 10:1111-1115(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Entian K.-D., Koelter P., Rose M., Li Z., Thermann R., Brendel M.,
RA Baur A., Boles E., Miosga T., Schaaff-Gerstenschlaeger I.,
RA Zimmermann F.R.;
RA Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
RL -1 SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1 SIMILARITY: TO YEAST YJL186W.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U05211; AAA21860.1; -
DR EMBL; Z35884; CAA84957.1; -.
DR PIR; S45870; S45870.
DR SGD; S0000219; MN2.
DR GO; GO:0005794; C:Golgi apparatus; IDA.
DR GO; GO:0000026; P:alpha-1,2-mannosyltransferase activity; IMP.
DR GO; GO:0006486; P:protein amino acid glycosylation; IMP.
KW Transmembrane; Glycoprotein; Signal-anchor.
FT DOMAIN 1 12
FT TRANSMEM 13 28
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 29 597
FT CARBOHYD 34 34
FT CARBOHYD 363 363
FT CARBOHYD 473 473
FT CONFLICT 287 287
FT CONFLICT T->K (IN REF. 1).
SQ SEQUENCE 597 AA; 67773 MW; 6CAFA1F4692A261C CRC64;

Query Match 5.8%; Score 96; DB 1; Length 597;
Best Local Similarity 24.2%; Pred. No. 7.4;
Matches 44; Conservative 27; Mismatches 67; Indels 44; Gaps 6;

OY 40 LLLIAPRAENNMVVOQDEYLESEFCMATRKISVITIFGPNVNSTMTKIDHFDLNEKPM 99
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 23 LEVITNKIWDENTSVKREKEYIDRVQOSTSNKSSSDAASADDSPLRDNDAGNERK-- 80

OY 100 RVVDEDDLDQRLISELRKEYGMTYNDFPMVLTVDLRVKQYEVPIPMKSVFDLIDTFQ 159
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 81 -----LKSFYNNVFN-FLMVDSPRGSTAKQYNACCLKGLGDGRPDHY- 122

OY 160 SRIKMEKOKKGGIVCKEYGVLELFPINGSSVVEREDVPAHLVKYDIRNFFQ-----VS 214
   | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 123 ---KDLVK-----LSAKLSKCLELSP-----DEVASLTKSHKDYVEHATLVS 163

OY 215 PE 216
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 164 PK 165

RESULT 10
HEP2_HUMAN STANDARD; PRT; 499 AA.
ID HEP2_HUMAN
AC P05546;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Heparin cofactor II precursor (HC-II) (Protease inhibitor leusepin 2)
DE (HLS2).

```


CN SERPINI1 OR HCF2.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91120782; PubMed=1671335;
 RA Herzog R., Lutz S., Blin N., Marasa J.C., Blinder M.A.,
 RT Tollefsen D.M.;
 RT "Complete nucleotide sequence of the gene for human heparin cofactor
 RT II and mapping to chromosomal band 22q11.";
 RL Biochemistry 30:1350-1357(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88163663; PubMed=2894851;
 RA Blinder M.A., Marasa J.C., Reynolds C.H., Deaven L.L., Tollefsen D.M
 RT "Heparin cofactor II: cDNA sequence, chromosome localization,
 RT restriction fragment length polymorphism, and expression in
 RT Escherichia coli.";
 RL Biochemistry 27:752-759(1988).
 RN [3]
 RP SEQUENCE OF 19-499 FROM N.A.
 RX MEDLINE=66120356; PubMed=3003690;
 RA Ragg H.;
 RT "A new member of the plasma protease inhibitor gene family.";
 RL Nucleic Acids Res. 14:1073-1088(1986).
 RN [4]
 RP SEQUENCE OF 333-499 FROM N.A.
 RX MEDLINE=86242236; PubMed=3755044;
 RA Inhorn R.C., Tollefsen D.M.;
 RT "Isolation and characterization of a partial cDNA clone for heparin
 RT cofactor III.";
 RL Biochem. Biophys. Res. Commun. 137:431-436(1986).
 RN [5]
 RP SEQUENCE OF 20-52 AND 464-499
 RX MEDLINE=66077723; PubMed=3907702;
 RA Griffith M.J., Noyes C.M., Tyndall J.A., Church F.C.;
 RT "Structural evidence for leucine at the reactive site of heparin
 RT cofactor II.";
 RL Biochemistry 24:6777-6782(1985).
 RN [6]
 RP SEQUENCE OF 1-119 FROM N.A.
 RX MEDLINE=88298901; PubMed=2841345;
 RA Ragg H., Preibisch G.;
 RT "Structure and expression of the gene coding for the human serpin
 RT hL52.";
 RL J. Biol. Chem. 263:12129-12134(1988).
 RN [7]
 RP SEQUENCE OF 58-85.
 RX MEDLINE=91093260; PubMed=1985958;
 RA Church F.C., Pratt C.W., Hoffman M.;
 RT "Leukocyte chemottractant peptides from the serpin heparin cofactor
 RT II.";
 RL J. Biol. Chem. 266:704-709(1991).
 RN [8]
 RP FUNCTION OF N-TERMINAL ACIDIC DOMAIN.
 RX MEDLINE=92041850; PubMed=1939083;
 RA van Deerlin V.M.D., Tollefsen D.M.;
 RT "The N-terminal acidic domain of heparin cofactor II mediates the
 RT inhibition of alpha-thrombin in the presence of glycosaminoglycans.";
 RL J. Biol. Chem. 266:20223-20231(1991).
 RN [9]
 RP MUTAGENESIS OF ARG-122 AND LYS-204.
 RX MEDLINE=90094412; PubMed=2104620;
 RA Blinder M.A., Tollefsen D.M.;
 RT "Site-directed mutagenesis of arginine 103 and lysine 185 in the
 RT proposed glycosaminoglycan-binding site of heparin cofactor II.";
 RL J. Biol. Chem. 265:286-291(1990).
 RN [10]
 RP VARIANT OSLO HIS-208.
 RX MEDLINE=89174798; PubMed=2647747;
 RA Blinder M.A., Andersson T.R., Abildgaard U., Tollefsen D.M.;

[illegible]

```

FT ACT_SITE 463 464 REACTIVE BOND (BY SIMILARITY).
FT VARIANT 7 7 A-> T (IN dbSNP:5905).
FT VARIANT 60 60 /FtId=VAR.011746.
FT VARIANT 60 60 H-> P (IN dbSNP:165867).
FT VARIANT 208 208 /FtId=VAR.011747.
FT VARIANT 208 208 R-> H (IN OSLO: HCF2 DEFICIENCY;
DECREASED AFFINITY FOR DERMATAN SULFATE;
dbSNP:6065).
FT VARIANT 237 237 /FtId=VAR.007112.
FT VARIANT 237 237 K-> R (IN dbSNP:1042435).
FT VARIANT 442 442 /FtId=VAR.011748.
FT VARIANT 442 442 T-> M (IN dbSNP:5904).
FT MUTAGEN 122 122 /FtId=VAR.011749.
FT MUTAGEN 122 122 R-> L: NORMAL THROMBIN INHIBITION AND
GLYCOSAMINOGLYCAN AFFINITY.
FT MUTAGEN 122 122 R-> O: GREATLY REDUCED THROMBIN
INHIBITION. NORMAL GLYCOSAMINOGLYCAN
AFFINITY.
FT MUTAGEN 122 122 R-> W: GREATLY REDUCED THROMBIN
INHIBITION. NORMAL GLYCOSAMINOGLYCAN
AFFINITY.
FT MUTAGEN 204 204 K-> M: REDUCED HEPARIN- AND NO DERMATAN
SULFATE-ACTIVATED INHIBITION.
FT MUTAGEN 204 204 K-> N: REDUCED HEPARIN- AND NO DERMATAN
SULFATE-ACTIVATED INHIBITION.
FT MUTAGEN 204 204 K-> T: REDUCED HEPARIN- AND NO DERMATAN
SULFATE-ACTIVATED INHIBITION.
FT CONFLICT 49 49 MISSING (IN REF. 5).
FT CONFLICT 483 483 R-> P (IN REF. 5).
FT CONFLICT 486 486 C-> T (IN REF. 5).
FT CONFLICT 499 499 S-> Q (IN REF. 5).
SQ SEQUENCE 499 AA; 57070 MW; 3B0E353FE1FDF05 CRC64;

Query Match 5.8%; Score 95; DB 1; Length 499;
Best Local Similarity 19.5%; Pred. No. 7;
Matches 70; Conservative 63; Mismatches 126; Indels 100; Gaps 18;

```

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RL *50 million years of genomic stasis in endosymbiotic bacteria.*;
Science 296:2376-2379(2002).
CC -1- FUNCTION: PHASES ACCELERATE THE FOLDING OF PROTEINS (BY
SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Peptidylproline (omega=180) - peptidylproline
(omega=0).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE P1IC/PARVULIN FAMILY OF ROTAMASES.
CC STRONG, TO E.COLI PPID.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE014122; AAM68005.1; -
CC Interpro: IPR000297; Rotamase.
CC Pfam: PF00639; Rotamase.1.
CC PROSITE: PS01096; P1IC_P1PIASE_1; FALSE_NEG.
CC PROSITE: PS01098; P1IC_P1PIASE_2; 1.
CC KW Isomerase; Rotamase; Transmembrane; Complete proteome.
CC FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 11 31 POTENTIAL.
CC FT DOMAIN 32 621 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 228 355 P1IC.
CC FT SEQUENCE 621 AA; 74139 MW; 5B7A89B253144C7C CRC64;

Query Match 5.7%; Score 93; DB 1; Length 621;
Best Local Similarity 19.7%; Pred. No. 13;
Matches 57; Conservative 47; Mismatches 97; Indels 88; Gaps 14;

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagsen K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrett B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT XIII";
 RL Nature 387:90-93(1997).
 RN [2]
 RP IDENTIFICATION OF INTRON.
 RA Rodriguez-Navarro S., Perez-Otin J.;
 RL Unpublished observations (xxx-1999).
 CC -1- SIMILARITY: TO YEAST YDR458C.
 CC -1- SIMILARITY: SOME, TO S.POMBE SPAC14C4.05C.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to the
 CC fact that an intron was not detected.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; Z46659; CAAB6621.1; ALT_SEQ.
 DR EMBL; Z46659; CAAB6622.1; ALT_SEQ.
 DR SGD; S0004497; SRCL.
 DR GO; GO:0016359; P:mitotic sister chromatid separation; ISI.
 DR Transmembrane.
 FT DOMAIN 117 129 POLY-ASP.
 FT TRANSMEM 455 475 POTENTIAL.
 FT SEQUENCE 834 AA; 95497 MW; 214CA064ECE160F4 CRC64;
 SQ
 Query Match 5.6%; Score 92.5; DB 1; Length 834;
 Best Local Similarity 17.6%; Pred. NO. 21;
 Matches 62; Conservative 71; Mismatches 105; Indels 115; Gaps 19;
 OY 30 LLSGFEGRRLRL-----ITAPKAKNNM--YVQORDYLRFSF----- 64
 DB 470 LFGIMYRQRLITGCGHEVSHRYSNGSFEIQLDMLDIDYRKCIICPPNGICPYL 529
 OY 65 --C--KMATRKISVITTFGPVNNSTMKIDHFDLNKEMPRVVDDELVDRLISELR 117
 DB 530 KLKCKPKDKLAPSRDLFEIT-PAQGKCVK-----DDKK-----QQLVSEV 570
 OY 118 K-----EKGMVYNDFEMLVITVDLRAVKQYEVPTIMKSVFDLI 155
 DB 571 EKSLFPLRAKNAQISCGDKDDISGDMEDALYQIFNFA-----RAPMIRDFEF- 620
 OY 156 DTFOSRIKMEKQKKEIVCE-----EVGVLLEFPIGSSVVEREDV--AHL--- 203
 DB 621 DLMIOVIRKDLFEFE--ILMROLSPDTNNIG-----NSNNIKITNDVVRQRKHLPEK 671
 OY 204 -VKDIRNFQVSPFEFSMLVGKQGNVKS--VYSPMKS--VIYVDLIDSMQLR--RQ 255
 DB 672 FISKTRNFRSTSKYIKGKCRFEIRIYQYKKFQPIIMLMLLITISKVIETKKNYRK 731
 OY 256 EMAIOOSLGMKQKMSQAMNTIVTKRTIRMTVITIMVITMDLTLEQY 308
 DB 732 KARIEIVTQIMERKLFO---KIRMSDPKENAVLSTIVOLDIFLSDIVDKY 781

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable translation initiation factor IF-2 (Contains: Pab infB
 DE Intron (Pab IF2 Intron)).
 GN INFB OR PYRAB11390 OR PAB0755.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GES / Orsay;
 RX PubMed=12622808;
 RA Cohen G.N., Barbe V., Tiamont D., Galperin M., Hellig R., Lecompte O.,
 RA Poch O., Priour D., Queffellon J., Ripp R., Thierry J.-C.,
 RA Van der Oost J., Weissbach J., Zivanovic Y., Forterre P.;
 RT "An integrated analysis of the genome of the hyperthermophilic
 RT archaeon Pyrococcus abyssi.";
 RL Mol. Microbiol. 47:1495-1512(2003).
 CC -1- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
 CC THE BINDING OF THE FORMYL METHIONINE-TRNA TO RIBOSOMES. SEEMS TO
 CC FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
 CC -1- PIM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
 CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
 CC -1- MISCELLANEOUS: THE INTEIN INTERRUPTS THE GTP BINDING SITE.
 CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
 CC -1- SIMILARITY: IN THE INTEIN SECTION, BELONGS TO THE HOMING
 CC ENDOUCLEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; AJ248286; CAB50050.1; -
 DR PIR; E75093; E75093.
 DR HAMAP; MF_00100; -, 1.
 DR InterPro; IPR000795; EF_GRPbind.
 DR InterPro; IPR004161; EFPU_D2.
 DR InterPro; IPR003586; Hedgehog_hintc.
 DR InterPro; IPR000178; Hedgehog_hintn.
 DR InterPro; IPR006141; Intron.
 DR InterPro; IPR004042; Intron_endonuc.
 DR InterPro; IPR005225; Small_GTP.
 DR InterPro; IPR004544; TIF_atf-2.
 DR Pfam; PF00009; GTP_EFTU; 1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR ProDom; PD186100; IF2; 1.
 DR SMART; SM00305; Hintc; 1.
 DR SMART; SM00306; Hintn; 1.
 DR TIGRFAMS; TIGR00491; atf-2; 1.
 DR TIGRFAMS; TIGR01443; Intron_Cterm; 1.
 DR TIGRFAMS; TIGR01445; Intron_Nterm; 1.
 DR TIGRFAMS; TIGR00231; small_GTP; 1.
 DR PROSITE; PS01176; IF2; FALSE_NEG.
 DR PROSITE; PS50818; INTEIN_C_TERM; 1.
 DR PROSITE; PS50819; INTEIN_ENDONUCLEASE; 1.
 DR PROSITE; PS50817; INTEIN_N_TERM; 1.
 DR Initiation factor; protein biosynthesis; GTP-binding;
 DR Autocatalytic cleavage; protein splicing; Hydrolase; Nuclease;
 DR Endonuclease; Intron homing; Complete proteome.
 FT CHAIN 1 20 PROBABLE TRANSLATION INITIATION FACTOR
 FT IF-2, 1ST PART (POTENTIAL).
 FT PAB INFB INTEIN (POTENTIAL).
 FT PROBABLE TRANSLATION INITIATION FACTOR
 FT IF-2, 2ND PART (POTENTIAL).
 FT GTP (BY SIMILARITY).
 FT NP_BIND 472 476

```

FT NP_BIND 526 529 GTP (BY SIMILARITY).
SQ SEQUENCE 992 AA: 112224 MW: 9838BADA50E6FIC6 CRC64;
Query Match
Best Local Similarity 5.6%; Score 92; DB 1; Length 992;
Matches 41; Conservative 36; Mismatches 73; Indels 24; Gaps 7;
OY 42 LITPAKANNMIVQORDEYLESFCK--MATKISVITTFGPVNNSTMKIDHFDLNEKPM 99
Db 720 VIAAPTPED--VEKAKOEILQIEERVISTDKVGIVRADLGSLEALSKEIQ--EKEIPI 776
OY 100 RVVD-----DEDIVDQRLISELRKEXGMTYNDFFVLDVLRVQYVEPTMKSV-FD 153
Db 777 RKADGVNSTKTDVMEALSVKEEPKYGVLGNVKNEDAE-EVAKADVKLEFVGNVLYK 835
OY 154 LIDFQSRKIMKMEKQKEGIVCKEEVGLLEFP-----INGSSVVE 195
Db 836 LIDEEVWKEEKKRELLSKVTFPGVIRLYPERKRVFRNSNATVGIETIE 889
RESULT 14
PIBL_BOVIN STANDARD; PRT: 1216 AA.
ID P10894;
AC P10894;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-Phosphatidylinositol-4,5-bisphosphate phospholipase beta 1
DE (EC 3.1.4.11) (Phosphoinositide phospholipase C) (PLC-beta-1)
DE (Phospholipase C-beta-1) (PLC-1) (PLC-154).
DE PLCB1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=88270496; PubMed=2455601;
RA Katan M., Kriz R.W., Totty N., Philipp R., Meldrum E., Aldape R.A.,
RA Knopf J.L., Parker P.J.;
RT "Determination of the primary structure of PLC-154 demonstrates
RT diversity of phosphoinositide-specific phospholipase C activities.";
RL Cell 54:171-177(1988).
CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES
CC DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS
CC MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
CC C ENZYMES.
CC -1- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-
CC bisphosphate + H(2)O = D-myo-inositol 1,4,5-trisphosphate +
CC diacylglycerol.
CC -1- COFACTOR: Calcium.
CC -1- MISCELLANEOUS: THE RECEPTOR-MEDIATED ACTIVATION OF PLC-BETA 1 IS
CC MEDIATED BY TWO G-PROTEIN ALPHA SUBUNITS, ALPHA-O AND ALPHA-11.
CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS
CC OF PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.
CC -1- SIMILARITY: Contains 1 C2 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, J03137; AAA30702.1; -.
CC PIR, A28822; A28822.
CC HSSP, P10688; 1DX.
CC InterPro: IPR000008; C2.
CC InterPro: IPR001192; PI_PLC.
CC InterPro: IPR000909; PI_PLC_Xdom.
CC InterPro: IPR001711; PI_PLC_Y.

```

```

DR Pfam: PF00168; C2; 1.
DR Pfam: PF00388; PI-PLC-X; 1.
DR Pfam: PF00387; PI-PLC-Y; 1.
DR PRINTS: PR00390; PPHPLIPASEC.
DR ProDom: PD001202; PI_PLC_Y; 1.
DR SMART: SM00239; C2; 1.
DR SMART: SM00148; PLCX; 1.
DR SMART: SM00149; PLCY; 1.
DR PROSITE: PS50004; C2-DOMAIN_2; 1.
DR PROSITE: PS50007; PIPLC_X_DOMAIN; 1.
DR PROSITE: PS50008; PIPLC_Y_DOMAIN; 1.
DR Hydrolase: Lipid degradation; Transducer; Phosphorylation; Calcium.
FT DOMAIN 316 467 DOMAIN X.
FT DOMAIN 540 761 DOMAIN Y.
FT DOMAIN 663 761 C2 DOMAIN.
FT ACT_SITE 331 331 BY SIMILARITY.
FT ACT_SITE 378 378 BY SIMILARITY.
FT MOD_RES 887 887 PHOSPHORYLATION (BY PKC).
SQ SEQUENCE 1216 AA: 138714 MW: BEF80917F1B7ABB CRC64;
Query Match
Best Local Similarity 19.9%; Score 92; DB 1; Length 1216;
Matches 56; Conservative 52; Mismatches 93; Indels 80; Gaps 11;
OY 7 EODGYQKPTNKHFTQSPKSVADLIGSFEGKRRLLITPAENNMVYQORDEY----- 60
Db 915 QQSEVVKLQKKHY-----KEMKDLVKRHHKRTTDLI-----KEHTTKYNEIONDYLRRAA 965
OY 61 LESFCKMATRKISVITTFGPVNNSTMKIDHFDLNEKPMRVVDDDLVDLRLSELREY 120
Db 966 LEKTARKDNKRKSESSPDHY--STTEODLALAEVMTQKLVLDKDKQOQDL----- 1017
OY 121 GMTYNDFFVLDVLRVQY-----EVPITKSVGFLLIDTFQS-----RIKDM-EK 167
Db 1018 -----NLROEQYYSEKYQKREHKILLQGLTFVAAECNNQKLLKEICEK 1063
OY 168 OKKEGIVCKEEVGLLEFPINGSSVEREDVPAHLVDIRNYFOVSEPYFSMLVKGDG 227
Db 1064 EKEE--LTKRKDKRKQ--EKITEAKSKDKRQOMEKEETMIRSIOEVQYIKRL----- 1113
OY 228 NVKSWYSPMWSMIVYDLISMOLRQEMAIOGSLGRQ 268
Db 1114 -----EBAQSKROEKLVKKEIKRQ 1134
RESULT 15
RNS2_SOLTU STANDARD; PRT: 223 AA.
ID RNS2_SOLTU
AC 001796;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribonuclease S-2 precursor (EC 3.1.27.1) (Stylar glycoprotein 2)
DE (S2-RNase).
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91246122; PubMed=2038308;
RA Kautmann H., Salamini F., Thompson R.D.;
RT "Sequence variability and gene structure at the self-incompatibility
RT locus of Solanum tuberosum.";
RL Mol. Gen. Genet. 226:457-466(1991).
CC -1- FUNCTION: SELF-INCOMPATIBILITY (SI) IS THE INHERITED ABILITY OF
CC A FLOWERING PLANT TO PREVENT SELF-FERTILIZATION BY DISCRIMINATING
CC BETWEEN SELF AND NON-SELF POLLEN DURING POLLINATION. IN MANY
CC SPECIES OF THE SOLANACEAE, SELF-INCOMPATIBILITY IS CONTROLLED BY
CC THE SINGLE, MULTIALLELIC LOCUS S. THIS STYLAR GLYCOPROTEIN IS
CC ASSOCIATED WITH EXPRESSION OF SELF-INCOMPATIBILITY IN POTATO.
CC -1- CATALYTIC ACTIVITY: Two-stage endonucleolytic cleavage to

```

CC	nucleoside 3'-phosphates and 3'-phosphooligonucleotides with
CC	2',3'-cyclic phosphate intermediates.
CC	-1- SUBCELLULAR LOCATION: Extracellular.
CC	-1- TISSUE SPECIFICITY: PISTIL.
CC	-1- SIMILARITY: BELONGS TO THE RNASE T2 FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: X62727; CAA44600.1; -.
DR	PIR: S16007; S16007.
DR	InterPro: IPR001568; RNase-T2.
DR	PIRfam: PF00445; Ribonuclease_T2; 1.
DR	PROSITE: PS00530; RNASE_T2_1; 1.
DR	PROSITE: PS00531; RNASE_T2_2; 1.
KW	Hydrolase; Nuclease; Endonuclease; Glycoprotein; Signal.
FT	SIGNAL
FT	CHAIN 1 22
FT	DISULFID 71 119
FT	DISULFID 178 211
FT	ACT_SITE 55 55
FT	ACT_SITE 112 112
FT	ACT_SITE 116 116
FT	CARBOHYD 51 51
SO	SEQUENCE 223 AA; 26075 MW; 9BAFCF5D737FDE CRC64; (POTENTIAL).
QY	Query Match 5.6%; Score 91.5; DB 1; Length 223;
QY	Best Local Similarity 22.1%; Pred. No. 4.6;
QY	Matches 47; Conservative 30; Mismatches 49; Indels 87; Gaps 10;
Db	26 SVADLLSFEKRRLLITAPKAEANNMTVQORDLYLSFCKMATRKISVITIGPVANNST 85
Db	16 SLSPIDYEDFDPMQ--LVLTPWRSEC-----YPRGFCNRI-----PPNNFT 53
QY	86 MKIDHFDLNEKPRP-----VVDEED-----LVDOFLIS 114
Db	54 I---HGIMPPKKPRKRGQLCTSDYIKETPGSVLDALDHNIWIDLKREIREIGINDPLMK 110
QY	115 ELRKEYGK---TYNDFEPMVLTDVLRVKQYEVPIITKRSVFDLIDTFQS----- 160
Db	111 DQYKKGHTCCLPYRNO-----QYFLAMRLKEKFDLLTLTRHTGTPGTRHT 158
QY	161 --RIKDMEK--OKKEGIVKEEYGVLELPEPI 188
Db	159 FKKIQAIAIKYVTEVPDCKVCEANIQGVLELEI 191

Search completed: July 24, 2003, 12:52:25
Job time : 10.8459 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:46:04 ; Search time 15.7015 Seconds
(without alignments)
1959.945 Million cell updates/sec

Title: US-09-884-319a-4
Perfect score: 1645
Sequence: 1 KKGGKTEODGYOKPTNKHPT.....DTLFEQKYVTLDSASFSLCSC 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1261.5	76.7	949	JC7802	urp protein - mouse
2	102	6.2	839	B96998	hypothetical protein
3	101.5	6.2	353	A82405	hypothetical protein
4	100	6.1	489	H86711	4-alpha-glucanotri
5	99.5	6.0	1134	D75014	hypothetical protein
6	99	6.0	1044	H71049	translation initia
7	98	6.0	1018	E64451	type I restriction
8	97	5.9	218	E84335	hypothetical prote
9	97	5.9	728	B71009	hypothetical prote
10	96.5	5.9	1769	S53378	probable membrane
11	96	5.8	349	D69346	hypothetical prote
12	96	5.8	397	P90114	26S protease regul
13	96	5.8	434	C86865	hypothetical prote
14	96	5.8	480	B45600	asparaginase-rich bl
15	96	5.8	597	S45870	TTPI protein - yea
16	95.5	5.8	476	H84524	probable fatty acyl
17	95.5	5.8	980	E71606	hypothetical prote
18	95	5.8	499	A37924	heparin cofactor I
19	95	5.8	613	H90160	conserved cytochet
20	93.5	5.7	489	B84733	probable cytochrom
21	93.5	5.7	1997	F71607	DNA helicase II BR
22	93	5.7	516	G96949	probable membrane
23	93	5.7	703	T15234	hypothetical prote
24	93	5.7	732	T32757	hypothetical prote
25	92.5	5.6	1166	T15628	hypothetical prote
26	92.5	5.6	1188	E89896	chromosome segrega
27	92.5	5.6	4307	T20721	hypothetical prote
28	92	5.6	992	E75093	translation initia
29	92	5.6	1216	A28622	1-phosphatidylinos

30	91.5	5.6	223	2	S16007	S-associated major
31	91.5	5.6	321	2	A71168	hypothetical prote
32	91.5	5.6	409	2	H89101	protein P25F5.5 (l
33	91.5	5.6	469	2	A44841	low molecularweig
34	91.5	5.6	502	1	D64110	lysine-tRNA ligase
35	91.5	5.6	761	2	E82205	chemotaxis protein
36	91.5	5.6	913	2	T33176	hypothetical prote
37	91	5.5	261	2	T46339	hypothetical prote
38	90.5	5.5	336	2	B64375	hypothetical prote
39	90.5	5.5	336	2	F81348	ABC transport syst
40	90.5	5.5	938	2	AB1530	transcription regu
41	90.5	5.5	1113	2	H84105	hypothetical prote
42	90.5	5.5	1403	2	H97131	uncharacterized, p
43	90.5	5.5	2207	2	T42759	Munc13-3 protein -
44	90	5.5	192	2	S32042	GTP-binding protei
45	90	5.5	901	2	E84210	DNA polymerase B1

ALIGNMENTS

RESULT 1

urp protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 03-Jun-2002
C:Accession: JC7802
R:Aoki, K.; Sun, Y.; Aoki, S.; Wada, K.; Wada, E.
Biochem. Biophys. Res. Commun. 290, 1282-1288, 2002
A:Title: Cloning, expression, and mapping of a gene that is upregulated in adipose ti
A:Reference number: JC7802; PMID:11812002; MUID:21670972
A:Accession: JC7802
A:Molecule type: mRNA
A:Residues: 1949 <NOK>
A:Cross-references: DDBJ:AB075019
A:Note: Three Arg codons are present in this ORF, the first Met residue is designated
C:Comment: This protein, a secretory protein, whose expression is tissue-specific, pl
C:Gene: urp
A:Map position: 16
C:Keywords: adipose tissue
F:1-23/Domain: highly hydrophobic, signal sequence #status predicted <SIG>
F:282-608/Region: lysine-rich #status predicted
F:523-524/Region: cleavage recognition site by furin (Arg-Arg) #status predicted

Query Match	Score	1261.5	DB 2:	Length	949:
Best Local Similarity	76.7%		Pred. No. 1.6e-83;		
Matches 250; Conservative	6;	Mismatches 11;	Indels 61;	Gaps 1;	
QY	1	KKGGKTEODGYOKPTNKHPTQSPKKSVDLLGSFEGKRRLLITAPKAEENMYVOORDXY	60		
DB	585	KKAGKTEODDNQKPTKHLAPSPKKSVDLLGSFEGKRRLLITPKAENMYVOORDXY	644		
QY	61	LESFCKMATRKISVITIFGPNVNSTMKIDHFDLDNEKPMRVVDDDLVDQRLISLKEY	120		
DB	645	LESFCKMATRKISVITIFGPNVNSTMKIDHFDLDNEKPMRVVDDDLVDQRLISLKEY	704		
QY	121	GMYNDPEFVLDVLDLRVQYEVPTTMSVPLDITPSSRIKDMKKKEIVCK----	176		
DB	705	GMYNDPEFVLDVLDLRVQYEVPTTMSVPLDITPSSRIKDMKKKEIVCK----	176		
QY	177	-----EEV	179		
DB	765	QSELENLFSFRMRRLVTSAPNDEWASQSLANGACNFGRLHITILLGVGEV	824		
QY	180	GGVLELFPINGSSVVEREDVPAHLVNDIRNYQVSEYPSMLVKGDKGVNKMYPSPMWS	239		
DB	825	GGVLELFPINGSSVVEREDVPAHLVNDIRNYQVSEYPSMLVKGDKGVNKMYPSPMWS	884		
QY	240	MYIVYDLTSMQLRQEMAIQOSLGMRC	267		
DB	885	MYIVYDLTSMQLRQEMAIQOSLGMRC	912		

RESULT 2
B96998
phosphoenolpyruvate synthase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B96998
R:Rolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium*
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B96998
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-839 <KUP>
A:Cross-references: GB:AF001437; PTDN:AAK78773.1; PTD:Q15023685; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
A:Genetics:
A:Gene: CAC0797

	Query Match	6.2%	Score 102;	DB 2;	Length 839;
	Best Local Similarity	19.3%	Pred. No. 14;		
	Matches	42;	Conservative	44;	Mismatches 86; Indels 46; Gaps 6;
QY	115 ELRREYGMTYNDFEMLVLTVDLRAVKQYE-----VPIITKMSVEDLI 155				
Db	23 EMRKE-GFRVVDGVIDISNTYKEITISNEKEDIKNILSTINKSNIDVLSIKLASIFDDF 81				
QY	156 DTPGSRKIDMKOKKEGEGIVCEKEEVGLVEFPINGSSVVEHEEDVPAHLVKQIRARVPOVSP 215				
Db	82 VINDSLVNEIDKRLAKGV-----KYAVSSGSIKE-----DLDN-LSEFAG 119				
QY	216 EYFMSLLVGGDNVYK---SWYSPAMNSMVIIVYDLIDMSQLRQDMAIQDLSLGRQKMS 271				
Db	120 QYSFNLINGIEEIKKAIIDCYKSMYTKGVLSTFDLNNLEVRLEMAVIYQEWQSEKSG 179				
QY	272 MQAMVATVITTDTRAVTRRTVIMKRVITMDLTLEQKRV 309				
Db	180 VAFVNPPLTGDIKEMVEVEEGLEAIVSGQVVEERTI 217				

RESULT 3
 A:82405
 hypochemical protein VCA0884 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
 C:Species: *Vibrio cholerae*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: A82405
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Claydon, R.A.; Gwin, M.L.; Dodson, R.J.;
 charadson, D.; Ermolaeva, M.D.; Vamathavan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, F.
 l., R.R.; Melakanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A:Reference number: A82055; MUID:20406833; PMID:10952301
 A:Accession: A82405
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-353 <HEP>
 A:Cross-references: GB:AE004416; GB:AE003853; NID:g9658312; PIDN:AAF96782.1; GSPDB:GN0001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Map: VCA0884
 A:Map position: 2

```

Query Match      6.2%  Score 101.5;  DB 2;  Length 353;
Best Local Similarity 19.3%  Pred. No. 5.1;
Matches 61;  Conservative 60;  Mismatches 140;  Indels 55;  Gaps 8;

QY      16  N K H T Q S K K K V A D L L S F E G K R L L I T A P K A E N N M V Q O R D E Y L S E F C M A T R K I S Y I 75
      11  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      66  M L Y S N D I M D V S I A N T F K D S - N V Q I L T L F K A E T R D T A Q L O O L S L ----- T L 115
      76  T I F G P V N N S T A K I H F Q L D N E K P M K V Y D D E D L V D O R L I S E L R K E Y G M T Y N D F M V L D V D 135

```

Db 116 TLQNPQLSLTNSIKRYDEGLNMAHQVEDAHNLTQQTIAQIQOEBVSIOAEIITNAQID 175

Qy 136 LARKQYVEVPITMKSVFDLIDPEOSRIDMEKQKEGI-----VKKEVGSVLEFP 187

Db 176 LMKQO-----IAAFKTAIAVQSORKKGLEFETIFGVVLAPITLGSLLIAG 221

Qy 188 INGSVVEREDVPAHLVKDIDINRYFOVSPEYSFMLLVGKGVNKSWMYDSPMKSNIYVDLI 247

Db 222 FGVSSTVEAQSSEISSLQSDIOS-----SLNTINHDQOTLSODDOQIINSLNALLSV 272

Qy 248 DSNQULROEAMIOIQSIGMRCKSMQAMVITYTDTFMYRMT-----TVIMRVITMD 301

Db 273 D--QVNDCALSRSLD-----TLQTYLSLYNNTNNVSNLTKAQDSQAVLIEQVWYQ 324

Qy 302 TLTEQKVTITLDSASF 317

Db 325 SAYNEWDILLEVASTL 340

RESULT 4
H86711
4.alpha.-glucanotransferase (EC 2.4.1.25) [Imported] - Lactococcus lactis subsp. lactis
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86711
R:Bohlin, A.; Winkler, P.; Manger, S.; Tallon, O.; Malarne, K.; Weissenbach, J.; Ehr
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: H86711
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-489 <STO>
A:Cross-references: GB:AE005176; PID:g12723605; PIDN:AAK04794.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: malO
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match	6.1%;	Score 100;	DB 2;	Length 489;
Best Local Similarity	18.1%;	Pred No. 10;		
Matches	67;	Conservative	56;	Mismatches 104; Indels 144; Gaps 16
QY	21	QSPKRSVADLLGSEFGKRRLLLTTPAKENNNMYVOQRDLEYLESFCKMATRKISVITTFGP	80	
		: : : : : :		
Db	139	QMPEDSVAGLLGQFS---PALMSAEETISXGIFLRQMGTERFINPFKIDWVIDEIFGR	195	
QY	81	VN-----NSTMKIDHPOLD-----NEKMRVYDDEP	106	
		: : : : :		
Db	196	DNRDMITQFLDYIGNGNTYFONEYNNOKKYKAEQLEWREGLYLKHENITLLKDDENP	255	
QY	107	-----LVDPRLISELRKEY-----GMTYNDEFVAILTDVDRVQXYEVPITMKRSVPDL	154	
		: : : : : : :		
Db	256	EKYHPRISLIQTISREFGDDYKGRLEKLYNDYFYGRNVEFKKAYELRPV-----	307	
QY	155	IDTFOSRIKDMEKQKKEGIVCKEENGV-----LELFPINGSSVYERE	197	
		: : : : :		
Db	308	-----LKDATNM-----LAGEEDGMYVPANVPDYVMNLILRLIETERPSPNRRVSPLN	356	
QY	198	DVP-----AHLVKDIRNYQVSPRE-----YFSMLVGKDGNVKSMY-PSPMM	238	
		: : : : : : :		
Db	357	EVPYLSVLTSSHDITSPLRAMWEEHNEELQRIYN-----EVMGWGEALYYASAEII	408	
QY	239	-----SMVIVYDLID-----SMOLRQEMAIOQ-----SLGMRQCKMSM	272	
		: : : : : : :		
Db	409	QEIYKRHLNSNMMAYILPIQDWLANSSEQLRKEDAKSEQINIPANPHYWNYRILHCO----	464	
QY	273	QAMVIVITTKD 283		
		: : : :		
Db	465	--LEFTLINOD 473		

RESULT 5

D75014

hypothetical protein PAB1257 - *Pyrococcus abyssi* (strain Orsay)C:Species: *Pyrococcus abyssi*

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

C:Accession: D75014

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: D75014

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1134 <KAM>

A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CA850554.1; PID:e151645

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1257

Query Match 6.0%; Score 99.5; DB 2; Length 1134;

Best Local Similarity 21.9%; Pred. No. 32;

Matches 37; Conservative 42; Mismatches 53; Indels 37; Gaps 8;

Db 104 DEDVDQRLSELREKGYNDPFWLTDVLRVQYVEPTMKSVFDLIDTFOSRK 163

Db 820 DKLTVNQKLSQLEKIKARIENN---IRDRLIDEXSEIKRMKSQFDKLD--KALE 872

Db 164 DMK--OKKE-----GIVCKEYGVGVLELF-PINGSSVEREDVPAHLVYDI 207

Db 873 EKEGSRKSEYKIIYDDPPTGVGVORREGANLAKVSELSNLOE-----AEKL 926

Db 208 RNYFQ-----VSPEFSMLVKGDNVKSWPSPMWSMVIYDLDISM 250

Db 927 RNYLELISRLTSPYSGIIVPSKE--VYEIKPVPKMEVNRVYTTINTY 973

RESULT 6

H71049

translation initiation factor aIF-2 PH1095 [similarity] - *Pyrococcus horikoshii*C:Species: *Pyrococcus horikoshii*

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 28-Jul-2000

C:Accession: H71049

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kuslida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: H71049

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1044 <KAM>

A:Cross-references: GB:AP000005; NID:g3236132; PIDN:BA430194.1; PID:d1031137; PID:g32575

A:Experimental source: strain 073

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1095

C:Superfamily: translation initiation factor aIF-2 PH1095; translation elongation factor

C:Keywords: GTP binding

F:434-581/Domain: translation elongation factor Tu homology <ETU>

Query Match 6.0%; Score 99; DB 2; Length 1044;

Best Local Similarity 25.5%; Pred. No. 31;

Matches 40; Conservative 38; Mismatches 61; Indels 18; Gaps 8;

Db 42 LITPAKENMYVOORDE---YLSFCMATRKISVITFSPVNNSTMKIDHFDLDNEK 97

Db 772 VTAAPTEPD---VERAKEEIMROIISVV-ISTDKGVIVKADTLGSLALSKELO-EKEI 826

Db 98 PKRVVD---DEDVDQRLSELREKGYNDPFWLTDVLRVQYVEPTMKSV- 150

Db 827 PIRKADVGNISTVDMEALSVKENPKYGVILG-FNVKVNDEAKVAKAKEVPLFVGN 884

Db 151 VEDLIDTFOSRKIDMEKOKKEGIVCKEEVGVLELFP 187

Db 885 IYKLIEDYEAWIKKEEKKRRLAKVFPFGYIKLXP 921

RESULT 7

E64451

type I restriction enzyme homolog - *Methanococcus jannaschii*C:Species: *Methanococcus jannaschii*

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Dec-2002

C:Accession: E64451

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Kleen, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannasc*

A:Reference number: A64300; MUID:96337999; PMID:8688087

A:Accession: E64451

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1018 <BDU>

A:Cross-references: GB:U67562; GB:L77117; NID:g1591830; PID:g1591843; TIGR:MJ214; PI

A:Map position: REV1159588-1156532

A:Start codon: GTG

C:Superfamily: type I site-specific deoxyribonuclease chain head

Query Match 6.0%; Score 98; DB 2; Length 1018;

Best Local Similarity 17.7%; Pred. No. 36;

Matches 66; Conservative 75; Mismatches 123; Indels 108; Gaps 18;

Db 16 NKHFQSPKK--SVADLGSFEGKRRLILTPKAKENMYVOORDEYLESFCKMATRK- 72

Db 333 NAFKDCFKINSVEELKVLVDIK-----ESNNENISEKGVYLVMAHKEDKRLK 384

Db 73 SVYTFGYPVNNSTMKIDHFDLDNEKPMRVYDDEDLVDQRLSEL- 117

Db 385 DFIESFG-----SIDKELLIRDEAHNTESGKFTLNKILKNAIAGFTGT 432

Db 118 -----KEY-----GMTYNDPFWLTDVLRVQYVEPTMKSV--FDLIDTFO 160

Db 433 PVHKRDMSTFKYVAPQEEFYLDLRF--TEESIKSEFTPLIMRVVAPKPDIDISE 488

Db 161 RIKDM-----EKOKKEGIVCKEEVGVLELFP--NSSVVEED--VPALVYDIRNY 210

Db 489 EIKNITEKLVDEADKIVSVSKETAETIKLSDLSKSSIKESKASYAETHLEDTEPF 548

Db 211 PQVSPEYFSLVGVGD-----GNVKSWPSPMWSMVIY-YDLIDSMOLR 253

Db 549 -----KFKAMVYVADDRKSCILFKKYLDLELKEIKINY--NENWQVYITTYIHNDVELE 600

Db 254 ROEMAIQOSLGRCKRMSMQ-----AMVTVTKTRMTVTRMTVIMRVITMDTL 303

Db 601 NKKELEKKGKGVNLEKMKWTEDFINKENPKILVYKK--LLTGFDAPILIKTYIHQF 657

Db 304 TEOKYVTLDSAS 315

Db 658 LKD-YLLQASA 668

Query Match 6.0%; Score 98; DB 2; Length 1018;

Best Local Similarity 17.7%; Pred. No. 36;

Matches 66; Conservative 75; Mismatches 123; Indels 108; Gaps 18;

Db 16 NKHFQSPKK--SVADLGSFEGKRRLILTPKAKENMYVOORDEYLESFCKMATRK- 72

Db 333 NAFKDCFKINSVEELKVLVDIK-----ESNNENISEKGVYLVMAHKEDKRLK 384

Db 73 SVYTFGYPVNNSTMKIDHFDLDNEKPMRVYDDEDLVDQRLSEL- 117

Db 385 DFIESFG-----SIDKELLIRDEAHNTESGKFTLNKILKNAIAGFTGT 432

Db 118 -----KEY-----GMTYNDPFWLTDVLRVQYVEPTMKSV--FDLIDTFO 160

Db 433 PVHKRDMSTFKYVAPQEEFYLDLRF--TEESIKSEFTPLIMRVVAPKPDIDISE 488

Db 161 RIKDM-----EKOKKEGIVCKEEVGVLELFP--NSSVVEED--VPALVYDIRNY 210

Db 489 EIKNITEKLVDEADKIVSVSKETAETIKLSDLSKSSIKESKASYAETHLEDTEPF 548

Db 211 PQVSPEYFSLVGVGD-----GNVKSWPSPMWSMVIY-YDLIDSMOLR 253

Db 549 -----KFKAMVYVADDRKSCILFKKYLDLELKEIKINY--NENWQVYITTYIHNDVELE 600

Db 254 ROEMAIQOSLGRCKRMSMQ-----AMVTVTKTRMTVTRMTVIMRVITMDTL 303

Db 601 NKKELEKKGKGVNLEKMKWTEDFINKENPKILVYKK--LLTGFDAPILIKTYIHQF 657

Db 304 TEOKYVTLDSAS 315

Db 658 LKD-YLLQASA 668

Db 658 LKD-YLLQASA 668

A:Accession: E84335
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-218 <STO>
A:Cross-references: GB:AE004437; NID:g10581289; PIDN:AAG20049.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1843C

Query Match
Best Local Similarity 23.1%; Score 97; DB 2; Length 218;
Matches 40; Conservative 29; Mismatches 58; Indels 46; Gaps 5;

QY 26 SVADLGSFGRRLILITAPKANNMYQQRDEILESCMATRKISVITTFGPVNNST 85
DB 5 ALLDLGN-ENRRRIQLARPK---CYTEISEIGVSPKAV----- 43
QY 86 MKIDHFDLNEKPMRVDDLDVDRILSELREKMGTVN-----DEFVY 130
DB 44 --IDHLQ-----QLDDAGIVESRVDQRRKRYHIAENLLEVOLSPFGYGAASAIPA 93
QY 131 LTVDLRVKQYVEPITMKSVFDLIDTFQSRIRKDEKQKEGIVCKEYGVYL 183
DB 94 STDDLRLARQYVSRIEERDSNAEYVELAKLKEINLEIRELSMAQRWYQGL 146

RESULT 9

B71009
hypothetical protein PH1368 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 28-Jul-2000
C:Accession: B71009
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekita, M.; Ohikubo, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuhida, N.; Oguchi, DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71009; MUID:98344137; PMID:9679194
A:Accession: B71009
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-728 <KAN>
A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30474.1; PID:g3257791
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1368
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB2446

Query Match
Best Local Similarity 20.0%; Score 97; DB 2; Length 728;
Matches 70; Conservative 55; Mismatches 115; Indels 110; Gaps 18;

QY 31 LGSFGRRLILITAPKANNMYQQRDEILESCMATRKISVITTFGPVNNSTMKIDH 90
DB 89 IGSYNAKGLFIVASMSSEGEKHAVKGRYFEGF-----EVIY---VNSSVTL-- 134
QY 91 FOLDNEKPMRVDDLDVDRILSELREKMGTYND---FFVYLDVLRVQO----- 140
DB 135 -----KRGITTEVPKDLI--LYGRYGIKRSNGTIFLYKLPSEVKKKEEREIYV 184
QY 141 --YFEV-----PI-----TMKSVFDLIDT-----FOS-RIKDEKOKK 170
DB 185 YFSEVTEEVNSIMPLTFENNGTETVAVTFSVYNAKPKDMVKKRYQGIETIKLRLK 244
QY 171 EGI-----VCKEEVGVLLEFPINGSVVEREDVPAHLVKDIRNYFQV----- 213
DB 245 ESTITQLKAVLPSSASHVYKESING-----DISFYVNVQPKMEALKITAPILYQEA 297
QY 214 ---SEFYRSMIL-VKDGNAVSWYSPM-WSWYIYYDLDSMQ--LRQEMAI----- 259
DB 298 EAGSKYTFESVLLSAGEDSVISMDVREPRDMKACILVDGVAQEIYLRGETKRVNLIETI 357
QY 260 ---QOSLGRCKQKMSQAMVITVTKDTRVVTMTIVIMRVITMDLTBEK 307

DB 358 PRNASLGY-----YESKISFTVRKSNGEVINKETITLGVNIKYTKQOK 401

RESULT 10

S53378
probable membrane protein YJL109c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein J0808
C:Species: Saccharomyces cerevisiae
C:Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S53378; S56887; S57359
R:Rasmussen, S.W.

submitted to the EMBL Data Library, February 1995
A:Description: A 37.5 kb region of yeast chromosome X includes the SWE1, MEF2, GSH1 a

A:Reference number: S53378
A:Accession: S53378

A:Molecule type: DNA
A:Residues: 1-1769 <RAS>

A:Cross-references: EMBL:X85021; NID:g728698; PID:g728701
R:Rasmussen, S.W.

submitted to the Protein Sequence Database, September 1995
A:Reference number: S56876
A:Accession: S56887

A:Molecule type: DNA
A:Residues: 1-1769 <RAM>

A:Cross-references: EMBL:Z49384; NID:g1008292; PID:g1008293; MIPS:YJL109c
R:Rasmussen, S.W.

Yeast 11, 873-883, 1995
A:Title: A 37.5 kb region of yeast chromosome X includes the SWE1, MEF2, GSH1 and CSD

A:Reference number: S57357; MUID:96090136; PMID:7483851
A:Accession: S57359

A:Molecule type: DNA
A:Status: nucleic acid sequence not shown; translation not shown

A:Residues: 1-1169 <RAF>

A:Cross-references: EMBL:X85021; NID:g728698; PIDN:CA59385.1; PID:g728701
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

C:Genetics:
A:Cross-references: SGD:S0003645

A:Map position: 10L
C:Keywords: transmembrane protein

Query Match
Best Local Similarity 18.3%; Score 96.5; DB 2; Length 1769;
Matches 46; Conservative 54; Mismatches 107; Indels 45; Gaps 8;

QY 81 VNSTMKIDHFDLNEKPMRV---VDDLDVDRILSELREKMGTYNDFEVLTDVYL 136
DB 1140 INKITEETPDQYDVRNRLKLVYSVLDETSDKLRINREFETLLEGVLFINSVEL 1199
QY 137 RKQYVEPITMKSVFDLIDTFQSRIRKDEKQKEGIVCKEYGVLEFPING----- 190
DB 1200 TFS-----CITSQNEEASDS-ETSLSDHTTEIKE--ILFVYLGAVLQILPDEFVNAV 1251
QY 191 --SVYEREDVPAHLVKDIRNYFQVSPYFMSLLVKGQGNVSWYSPMMSWYIYYDLID 248
DB 1252 PLISTSTNEDIYHIL-----LVIGSFEELEGSATIPVNVNMYV-LLD 1294
QY 249 SMLRQEMALIQOSLGMRCQKMSQAMVITVTKDTRVVTMTIVIMRVITMDLTBEKY 308
DB 1295 RMPLEKSVYISQV-----LMTMTALVSKYKCKLEGSILTFQALTLATFEKVSQDME 1346
QY 309 VTLDSASFSCSC 320
DB 1347 VKISSIALITNC 1358

RESULT 11

D69346

hypothetical protein AF0772 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: D69346

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E

```

Query Match      5.88; Score 96; DB 2; Length 397;
Best Local Similarity 20.28; Pred. No. 15;
Matches 61; Conservative 61; Mismatches 126; Indels 54; Gaps 14;

QY      50 NNNVVQ-----QREYVLESFCCKMATRKISVITTFGPNNSTMTKIDHQLDNEKMRV-VD 103
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      34 NNNETIKLDMKKKEIQEIKSSSTLPHLVATISEIYYONNRKIDGILLKTSNRNLIFLS 93

QY      104 DEDVDQRRLISELKEKGMTYNDFMWL-----TDVDLAVK-----QYEVPTMTKS 150
      || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      94 NEGIDIKKNLN-LNSIGYVKKDSYLLISRLPEYDFEVRAMEVDENQGYTDIGLDNQ 152

QY      151 VFDDIDPFQSRIRKMEQKEGIVCKREVGVELEF--PLNGSSVVEREDVPALHLYDIR 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 14
B45600
asparagine-rich blood stage antigen (clone Pfafs-6) - Plasmodium falciparum (fragment
C:Species: Plasmodium falciparum
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: B45600; S27828
R:Notley, D.; Knapp, B.
Mol. Biochem. Parasitol. 46, 319-321, 1991
A:Title: Partial sequences of three new asparagine-rich blood stage proteins of Plas
A:Reference number: A45600; MUID:92018031; PMID:1922204
A:Accession: B45600
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-480 <NOL>

```

A:Cross-references: EMBL:M59472; NID:g160105; PIDN:AAA29497.1; PID:g552186
A:Experimental source: isolate FCBR

Query Match 5.8%; Score 96; DB 2; Length 480;

Best Local Similarity 21.8%; Pred. No. 19; Mismatches 77; Indels 72; Gaps 14;

Matches 53; Conservative 41; Mismatches 77; Indels 72; Gaps 14;

QY 81 VNNSTMKIDHFDLNEKPMRV-----VDEDEDLVDRQLISELRKEYGMTY-- 124

DB 86 VNEERVAVNDH---DNEEKNLVETQENENDKRNKKEDEDEVDYD---AEKEEKEGKMLLR 139

QY 125 -NDPFMVLTVDLARKYQYEPITMKSVFDLIDTQSRKIKMEKQK---EGT--VCKE 177

DB 140 EKICEKMLKSIDI-----YITEIIMKSGFVY---YRMKDELKWKRADIEGFLYIVKR 190

QY 178 EVGGLVLEFPINGSSVEREDVPAHLVDRIRNYFOVSPE--YFSMLLVGKDGNSWKVPS 235

DB 191 SIKPYRL-----ITNKKNEKHLDDIDTNNMLSTDONTYIFRIINEETNVRNRY-- 241

QY 236 PMMSWVIYVLDLSMQLRR-----QEMAIQSLGMRQCMQSMQAMVTVITTKDTPMVTM 290

DB 242 SLW---FYSTEKEQIYKVLKNIYKAYQEKVG-----ITQDNTNNSNM 283

QY 291 TTV 293

DB 284 NNI 286

RESULT 15

S45870

TPP1 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YBR015c; protein YBR0220

C:Species: Saccharomyces cerevisiae

C:Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 28-May-1999

C:Accession: S45870; S45595

R:Enitan, K.D.; Koetter, P.; Rose, M.; Li, Z.; Thermann, R.; Brendel, M.; Baur, A.; Bole

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45862

A:Accession: S45870

A:Molecule type: DNA

A:Residues: 1-597 <ENT>

A:Cross-references: EMBL:Z35884; NID:g536213; PID:g536214; MIPS:YBR015c

A:Experimental source: strain S288C

R:Romero, P.A.; Athanassiadis, A.; Lussier, M.; Herscovics, A.

Yeast 10, 1111-1115, 1994

A:Title: The nucleotide sequence of TPP1, a gene encoding a predicted type II membrane p

A:Reference number: S45595; MUID:95084638; PMID:7992511

A:Accession: S45595

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-286, 'K', 288-597 <ROM>

A:Cross-references: GB:U05211; NID:g500627; PIDN:AAA21860.1; PID:g500628

C:Genetics:

A:Gene: SGD:TPP1

A:Cross-references: SGD:S0000219; MIPS:YBR015c

A:Map position: 2R

C:Keywords: glycoprotein; transmembrane protein

F:13-28/Domain: transmembrane #status predicted <TM>

F:29-597/Domain: extracellular #status predicted <EXT>

F:34,363,473/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 5.8%; Score 96; DB 2; Length 597;

Best Local Similarity 24.2%; Pred. No. 25;

Matches 44; Conservative 27; Mismatches 67; Indels 44; Gaps 6;

QY 40 LLLITAPKAEENMYQOQDEYIESCKMATRKISVITTFGPVNNSTMKIDHFDLNEKPM 99

DB 23 LEVITNKMYDENTSVKKEYLDRYQSYSNKYSDDASADSTPLRDNDEAGNEK-- 80

QY 100 RVVDEDEDLVDRQLISELRKEYGMTYNDPFMVLTVDLARKYQYEPITMKSVFDLIDTQ 159

DB 81 -----LKSFTNNVFN-FLMVDSPKSGTAKQYNEACLKGDIGDRPDHY- 122

QY 160 SRIDMEKQKKEGIVCKEEVGVLEFPINGSSVEREDVPAHLVKDIRNYFO-----VS 214

DB 123 ---KDLVK-----LSAKELSKCLELSP-----DEVASLTKSHKDYEHIAVLVS 163

QY 215 PE 216

DB 164 PK 165

Search completed: July 24, 2003, 12:53:44

Job time : 18.7015 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:53:55 ; Search time 21.6724 Seconds
(without alignments)
1753.524 Million cell updates/sec

Title: US-09-884-319A-4
Perfect score: 1645
Sequence: 1 KKGGKTEODGYOKPTNKHFT.....DLTFEQKYTLDSASFICSC 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1645	100.0	320	9	US-09-840-707A-7
2	1645	100.0	320	11	US-09-884-319-4
3	1645	100.0	320	15	US-10-038-557A-7
4	1343.5	81.7	950	9	US-09-823-356-9
5	1057.5	64.3	316	9	US-09-925-301-947
6	921	56.0	260	15	US-10-074-475-263
7	513.5	31.2	189	10	US-09-764-877-1151
8	147.5	9.0	431	15	US-10-102-806-717
9	105.5	6.4	1182	15	US-10-156-634A-4
10	95.5	5.8	504	9	US-09-883-797-6
11	95	5.8	510	9	US-09-925-302-486
12	93.5	5.7	1184	9	US-09-815-242-5229
13	93.5	5.7	1188	9	US-09-815-242-12125
14	91.5	5.6	302	9	US-09-815-242-11205
15	89.5	5.4	1618	10	US-09-963-875-1

16	89.5	5.4	1618	15	US-10-136-891-2	Sequence 2, Appli
17	89.5	5.4	1618	15	US-10-120-687-1	Sequence 1, Appli
18	88.5	5.4	1145	15	US-10-116-949-2	Sequence 2, Appli
19	88.5	5.4	1145	15	US-10-116-949-4	Sequence 2, Appli
20	86.5	5.3	356	15	US-10-014-101-10	Sequence 35, Appli
21	86.5	5.3	539	15	US-10-014-101-35	Sequence 35, Appli
22	86.5	5.3	704	15	US-10-128-714-8238	Sequence 8238, Ap
23	86	5.2	494	15	US-10-043-487-296	Sequence 296, Ap
24	86	5.2	610	9	US-09-815-242-11183	Sequence 11183, A
25	86	5.2	723	15	US-10-083-357-1343	Sequence 1343, Ap
26	85.5	5.2	2710	15	US-10-011-366-6	Sequence 6, Appli
27	84.5	5.1	901	15	US-10-208-948-16	Sequence 16, Appli
28	84	5.1	2697	10	US-09-961-527A-5	Sequence 5, Appli
29	83.5	5.1	909	15	US-10-163-214-2	Sequence 2, Appli
30	83.5	5.1	1139	11	US-09-820-843A-15	Sequence 15, Appli
31	83.5	5.1	1462	15	US-10-128-714-8540	Sequence 8540, Ap
32	82.5	5.0	531	15	US-10-128-714-8198	Sequence 8198, Ap
33	82.5	5.0	608	10	US-09-738-626-3609	Sequence 3609, Ap
34	82	5.0	422	10	US-09-779-307-18	Sequence 18, Appli
35	82	5.0	482	10	US-09-779-307-17	Sequence 17, Appli
36	82	5.0	483	10	US-09-919-497-79	Sequence 79, Appli
37	82	5.0	483	10	US-09-974-298-41	Sequence 41, Appli
38	82	5.0	483	10	US-09-981-353-158	Sequence 158, App
39	82	5.0	483	15	US-10-177-293-258	Sequence 258, App
40	81.5	5.0	208	14	US-10-107-813-4	Sequence 4, Appli
41	81.5	5.0	415	10	US-09-925-300-1352	Sequence 1352, Ap
42	81.5	5.0	724	9	US-09-759-010-6	Sequence 6, Appli
43	81.5	5.0	724	11	US-09-932-257A-27	Sequence 27, Appli
44	81.5	5.0	1805	10	US-09-820-843A-73	Sequence 73, Appli
45	81	4.9	714	10	US-09-768-877-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1
US-09-840-707A-7
; Sequence 7, Application US/09840707A
; Patent No. US20020077276A1
; GENERAL INFORMATION:
; APPLICANT: Fredetking, Terry M.
; APPLICANT: Ignatyev, George M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
; TITLE OF INVENTION: INJECTIONS AND OTHER DISORDERS
; FILE REFERENCE: 24881-301C
; CURRENT APPLICATION NUMBER: US/09/840, 707A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/562, 979
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/198, 210
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: IL-1 receptor intracellular ligand protein
; OTHER INFORMATION: comprising amino acid sequence
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5, 817, 476
; PATENT FILING DATE: 1995-06-07
; PUBLICATION DATE: 1998-10-06
; US-09-840-707A-7

Query Match 100.0%; Score 1645; DB 9; Length 320;
Best Local Similarity 100.0%; Pred. No. 3.6e-154;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKGGKTEODGYOKPTNKHFTQSPKKSVDLLGSESKRRLLLTAPKAENMYVOORDEY 60
Db 1 KKGGKTEODGYOKPTNKHFTQSPKKSVDLLGSESKRRLLLTAPKAENMYVOORDEY 60

Qy	61	LESFCKMTRKISVTTIFCPVNNSTPKIDHFOLDNCKPMRVYDDEDLVDORLSLTKREY	120
Db	61	LESFCKMTRKISVTTIFCPVNNSTPKIDHFOLDNCKPMRVYDDEDLVDORLSLTKREY	120
Qy	121	GMTYDDEFMWLTDVDLRYKQYEVPIPTMKSVEDLIDTFOSRIDKMEKOKKEGIVCKEEVG	180
Db	121	GMTYDDEFMWLTDVDLRYKQYEVPIPTMKSVEDLIDTFOSRIDKMEKOKKEGIVCKEEVG	180
Qy	181	GYLELFPINGSVVEREDYPAHLVKDIRNFQVSPPEYFSLVLVGKGNKSWPSPMWSM	240
Db	181	GYLELFPINGSVVEREDYPAHLVKDIRNFQVSPPEYFSLVLVGKGNKSWPSPMWSM	240
Qy	241	VIVYDLDISMOLRQEMALQOISLGMRQKMSQOAMVTIVTTKDRVTRRTVIMRVITM	300
Db	241	VIVYDLDISMOLRQEMALQOISLGMRQKMSQOAMVTIVTTKDRVTRRTVIMRVITM	300
Qy	301	DTLTFQKRYVTLIDASFLICSC	320
Db	301	DTLTFQKRYVTLIDASFLICSC	320

RESULT 2

```

US-09-884-319-4
: Sequence 4, Application US/09884319
: Publication No. US20030124625A1
: GENERAL INFORMATION:
: APPLICANT: Lin, Lih-Ling
:              Graham, James
: TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
:                   INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
:                   BINDING
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
:   ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
:   STREET: 87 Cambridgepark Drive
:   CITY: Cambridge
:   STATE: MA
:   COUNTRY: USA
:   ZIP: 02140
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/09/884,319
:   FILING DATE: 18-Jun-2001
:   CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: US/09/083,516
:   FILING DATE: <Unknown>
:   APPLICATION NUMBER: 08/487,942
:   FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
:   NAME: Brown, Scott A.
:   REGISTRATION NUMBER: 32,724
:   REFERENCE/DOCKET NUMBER: G15258
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (617) 498-8224
:   TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 4:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 320 amino acids
:     type: amino acid
:     TOPOLOGY: linear
:     MOLECULE TYPE: protein
:     SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-884-319-4

```

Query Match	100.0%;	Score 1645;	DB 11;	Length 320;
Best Local Similarity	100.0%;	Pred. No. 3.6e-154;		
Matches 320;	Conservative	0;	Mismatches	0;
		Indels	0;	Gaps

QY	1	KKGKGTEDOGYOKPTNKHFTOSPKKSVADLLGSEPGKRRLLITAPAKENNNVYQOODEY	60
Db	1	KKGKGTEDOGYOKPTNKHFTOSPKKSVADLLGSEPGKRRLLITAPAKENNNVYQOODEY	60
QY	61	LESEFCMMATRKISVITIFGPNVNNSTMKIDHFPOLDNEKPMRVVDEDDLDVQRLISELREKEY	120
Db	61	LESEFCMMATRKISVITIFGPNVNNSTMKIDHFPOLDNEKPMRVVDEDDLDVQRLISELREKEY	120
QY	121	GMTYNDFFMWLTDVDLRAKYQYEVPTMKSVFDLLIDTFOSRIKIMEKOKKEGIYCKEEVG	180
Db	121	GMTYNDFFMWLTDVDLRAKYQYEVPTMKSVFDLLIDTFOSRIKIMEKOKKEGIYCKEEVG	180
QY	181	GVLELEPPINGSSVVEBEDYPAHLVYNDIRYRQVSPETYSMLLVKDGDNKWSNPSPPMSM	240
Db	181	GVLELEPPINGSSVVEBEDYPAHLVYNDIRYRQVSPETYSMLLVKDGDNKWSNPSPPMSM	240
QY	241	VIIYVDLIDDSQJLROBMAIQOISLGMRCKOMSQAOAVTITVTRKDTIRMYTBMITYMIRIYM	300
Db	241	VIIYVDLIDDSQJLROBMAIQOISLGMRCKOMSQAOAVTITVTRKDTIRMYTBMITYMIRIYM	300
QY	301	DTLEEQKYVTLDSASFCLSC	320
Db	301	DTLEEQKYVTLDSASFCLSC	320

RESULT 3

US-10-038-557A-7
Sequence 7, Application US/10038557A
Publication No. US20030092684A1
GENERAL INFORMATION:
APPLICANT: Friedeking, Terry M.
TITLE OF INVENTION: Inat'yev, George M.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
FILE REFERENCE: 24881-301D
CURRENT APPLICATION NUMBER: US/10/038,557A
CURRENT FILING DATE: 2002-06-05
PRIOR APPLICATION NUMBER: 09/840,707
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: 09/562,979
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/198,210
PRIOR FILING DATE: 1999-04-27
NUMBER OF SEQ ID NOS.: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 320
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: IL-1 receptor intracellular ligand protein
OTHER INFORMATION: comprising amino acid sequence
PUBLICATION INFORMATION:
PATENT DOCUMENT NUMBER: 5,811,476
PATENT FILING DATE: 1995-06-07
PUBLICATION DATE: 1998-10-06
US-10-038-557A-7

Query Match	100.0%;	Score 1645;	DB 15;	Length 320;
Best.Local Similarity	100.0%;	Pred. No. 3.6e-154;		
Matches 320;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 1 KKGKGTEDDGYQKPTNKHHTOSPCKSVADLIGSEFGKRRLLITAPRAENMYVQQRDEY 60

Dd 1 KKGKGTEDDGYQKPTNKHHTOSPCKSVADLIGSEFGKRRLLITAPRAENMYVQQRDEY 60

QY 61 LESFCCKMARRKISVITIFGPVNNSTMKIDHQLDNKRMRYVDEDEDLVDRQLSELKREY 120

Dd 61 LESFCCKMARRKISVITIFGPVNNSTMKIDHQLDNKRMRYVDEDEDLVDRQLSELKREY 120

QY 121 GMRYNDFPMVLVDVLRKQYREYVPIITKMSYVDLIDTPQSRKIMKMKQKKKGLVCKREY 180

Db 121 GMTYNDFEMLTVDLRAVKQYEVPIITMKSVFDLIDTFQSRIDMEKOKKEGIVCKEEVG 180
QY 181 GVLELFPINGSSVVEREDYPAHLVKDIRNYFOVSPPEYFMSMLLVGKGNVKSMPSPMMS 240
Db 181 GVLELFPINGSSVVEREDYPAHLVKDIRNYFOVSPPEYFMSMLLVGKGNVKSMPSPMMS 240
QY 241 VIVYDLIDSMQLRQEMAIQOISLGRCKMSQANVTITVTRKDRFMVTRMTVIMRVITM 300
Db 241 VIVYDLIDSMQLRQEMAIQOISLGRCKMSQANVTITVTRKDRFMVTRMTVIMRVITM 300
QY 301 DLTFRQKVTITLDSASFCLSC 320
Db 301 DLTFRQKVTITLDSASFCLSC 320

RESULT 4

US-09-823-356-9
; Sequence 9, Application US/09823356
; Patent No. US20010025098A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Kaaser, Matthew R.
; APPLICANT: Baughn, Marian R.
; APPLICANT: Shah, Puri
; TITLE OF INVENTION: HUMAN MEMBRANE SPANNING PROTEINS
; FILE REFERENCE: PF-0489-1 CON
; CURRENT APPLICATION NUMBER: US/09/823,356
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/039,307
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID NO. US20010025098A1 1794154
US-09-823-356-9

Query Match 81.7%; Score 1343.5; DB 9; Length 950;
Best Local Similarity 81.4%; Pred. No. 1e-123;
Matches 267; Conservative 0; Mismatches 0; Indels 61; Gaps 1;

QY 1 KKGATGTEODGYCKPINKHFTQSPKSVADLGSFGRRLILITAPKANNMYVOORDEY 60
Db 586 KKGATGTEODGYCKPINKHFTQSPKSVADLGSFGRRLILITAPKANNMYVOORDEY 645
QY 61 LESFCMATRKISVTITFGPVNNSMTMKIDHFDLDNEKPMRVVDDDELDVQRLISELRKEY 120
Db 646 LESFCMATRKISVTITFGPVNNSMTMKIDHFDLDNEKPMRVVDDDELDVQRLISELRKEY 705
QY 121 GMTYNDFEMLTVDLRAVKQYEVPIITMKSVFDLIDTFQSRIDMEKOKKEGIVCK --- 176
Db 706 GMTYNDFEMLTVDLRAVKQYEVPIITMKSVFDLIDTFQSRIDMEKOKKEGIVCKEDKK 765
QY 177 -----EYV 179
Db 766 QLENFSLSRFRRLVISPNDDEDMAYSQQLSALSGQCNFGLRHITILKLGVEEV 825
QY 180 GGVLELFPINGSSVVEREDYPAHLVKDIRNYFOVSPPEYFMSMLLVGKGNVKSMPSPMMS 239
Db 826 GGVLELFPINGSSVVEREDYPAHLVKDIRNYFOVSPPEYFMSMLLVGKGNVKSMPSPMMS 885
QY 240 MVIYVDLIDSMQLRQEMAIQOISLGRCK 267
|||||

Db 886 MVIYVDLIDSMQLRQEMAIQOISLGRCK 913.

RESULT 5

US-09-925-301-947
; Sequence 947, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 947
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (312)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-947

Query Match 64.3%; Score 1057.5; DB 9; Length 316;
Best Local Similarity 77.7%; Pred. No. 3.7e-96;
Matches 212; Conservative 0; Mismatches 0; Indels 61; Gaps 1;

QY 56 QDEYLESEFCMATRKISVTITFGPVNNSMTMKIDHFDLDNEKPMRVVDDDELDVQRLISE 115
Db 7 QDEYLESEFCMATRKISVTITFGPVNNSMTMKIDHFDLDNEKPMRVVDDDELDVQRLISE 66
QY 116 LRKEYGMTYNDFEMLTVDLRAVKQYEVPIITMKSVFDLIDTFQSRIDMEKOKKEGIVC 175
Db 67 LRKEYGMTYNDFEMLTVDLRAVKQYEVPIITMKSVFDLIDTFQSRIDMEKOKKEGIVC 126
QY 176 K----- 176
Db 127 KEDKQSLLENFLSRFRRLVISPNDDEDMAYSQQLSALSGQCNFGLRHITILKLG 186
QY 177 --EEYGVLELFPINGSSVVEREDYPAHLVKDIRNYFOVSPPEYFMSMLLVGKGNVKSMP 234
Db 187 VGEYGVLELFPINGSSVVEREDYPAHLVKDIRNYFOVSPPEYFMSMLLVGKGNVKSMP 246
QY 235 SPMSMVIYVDLIDSMQLRQEMAIQOISLGRCK 267
Db 247 SPMSMVIYVDLIDSMQLRQEMAIQOISLGRCK 279
|||||

RESULT 6

US-10-074-475-263
; Sequence 263, Application US/10074475
; Publication No. US20030092898A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Hervé
; APPLICANT: Kaira, Kalpana
; APPLICANT: Caffierkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0313

;; CURRENT APPLICATION NUMBER: US/10/074,475
;; CURRENT FILING DATE: 2002-02-13
;; PRIOR APPLICATION NUMBER: 60/268,292
;; PRIOR FILING DATE: 2001-02-13
;; NUMBER OF SEQ ID NOS: 295
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 263
;; LENGTH: 260
;; TYPE: PRT
;; ORGANISM: Homo sapien
US-10-074-475-263

Query Match 56.0%; Score 921; DB 15; Length 260;
Best Local Similarity 96.8%; Pred. No. 8.3e-83;
Matches 180; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 KKGKTEQDGYOKPTKHKHFTOSPKKSVADLLGSFECKRLLITAPKAENMYVOORDEX 60
DB 38 KKGKTEQDGYOKPTKHKHFTOSPKKSVADLLGSFECKRLLITAPKAENMYVOORDEX 97
OY 61 LESFCMATRKISVITIFGPVNNSTMKIDHFDLDNEKPMRVVDEDLVDQRLISELRKEY 120
DB 98 LESFCMATRKISVITIFGPVNNSTMKIDHFDLDNEKPMRVVDEDLVDQRLISELRKEY 157
OY 121 GMTYNDENVYLDVDLRVKQYEVPTMKSVFDLIDFOSRIKDEKQKKEGIVCKEYVG 180
DB 158 GMTYNDENVYLDVDLRVKQYEVPTMKSVFDLIDFOSRIKDEKQKKEGIVCKEDKK 217
OY 181 GVLELF 186
DB 218 GQLENF 223

RESULT 7
US-09-764-877-1151
;; Sequence 1151, Application US/09764877
;; Patent No. US20020147140A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC005
;; CURRENT APPLICATION NUMBER: US/09/764,877
;; CURRENT FILING DATE: 2001-01-17
;; Prior application data removed - refer to PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 4031
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1151
;; LENGTH: 189
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (14)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;; NAME/KEY: SITE
;; LOCATION: (15)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;; NAME/KEY: SITE
;; LOCATION: (183)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;; NAME/KEY: SITE
;; LOCATION: (189)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1151

Query Match 31.2%; Score 513.5; DB 10; Length 189;
Best Local Similarity 42.6%; Pred. No. 9.1e-43;
Matches 118; Conservative 12; Mismatches 30; Indels 117; Gaps 2;

OY 21 GSPKSVADLLGSFECKRLLITAPKAENMYVOORDEXLESFCMATRKISVITIFP 80
DB 23 EDKKSQDNFSLFRKRRRLVLSAPNDEDMAYSQQLSLSQACNFGRLHITLKLGG- 81

OY 81 VNNSTMKIDHFDLDNEKPMRVVDEDLVDQRLISELRKEYGMTYNDENVYLDVLRVKQ 140
DB 82 ----- 81
OY 141 YIEVPTMKSVFDLIDFOSRIKDEKQKKEGIVCKEYGVLELFPINGSSVVERDVP 200
DB 82 -----VGEYGVLELFPINGSSVVERDVP 107
OY 201 AHLVKDIRNYFOVSPFEYSMLLVGKGNKYSWYPSPMNSMIVYVLDISMQLRQEMAIQ 260
DB 108 AHLVKDIRNYFOVSPFEYSMLLVGKGNKYSWYPSPMNSMIVYVLDID----- 155
OY 261 QSLGMRQKMSQAMVTVTTKDTMRVTMTTVIMRV 297
DB 156 -----SMQAMVTVTTKDTMRVTMTTVIMRV 182

RESULT 8
US-10-102-806-717
;; Sequence 717, Application US/10102806
;; Publication No. US20030054421A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;; FILE REFERENCE: PA103PLCI
;; CURRENT APPLICATION NUMBER: US/10/102,806
;; CURRENT FILING DATE: 2002-03-22
;; PRIOR APPLICATION NUMBER: 09/925,298
;; PRIOR FILING DATE: 2001-08-10
;; PRIOR APPLICATION NUMBER: PCT/US00/05881
;; PRIOR FILING DATE: 2000-03-08
;; PRIOR APPLICATION NUMBER: 60/124,270
;; PRIOR FILING DATE: 1999-03-12
;; NUMBER OF SEQ ID NOS: 846
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 717
;; LENGTH: 431
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-102-806-717

Query Match 9.0%; Score 147.5; DB 15; Length 431;
Best Local Similarity 26.5%; Pred. No. 4.2e-06;
Matches 40; Conservative 29; Mismatches 67; Indels 15; Gaps 1;

OY 26 SVADLLGSFECKRLLITAPKAENMYVOORDEXLESFCMATRKISVITIFGPVNNST 85
DB 296 SAAGLLDQFEYKORLLIISAPDPSNRYYKMQISMLQOSTCGLDLRHVITIELVGQPPQEV 355
OY 86 MKIDHFDLDNEKPMRVVDEDLVDQRLISELRKEYGMTYNDENVYLDVLRVKQYEV 145
DB 356 GRIRGQL-----SANITEELNQFQRLRYSYENVYLDIKQIGDRQRYMEP 400
OY 146 ITMKSVFDLIDFOSRIKDEKQKKEGIVCK 176
DB 401 VTPEEFTFIDYLLSNQELTORREORDICE 431

RESULT 9
US-10-156-634A-4
;; Sequence 4, Application US/10156634A
;; Publication No. US20030013152A1
;; GENERAL INFORMATION:
;; APPLICANT: Delong, Jeff L.
;; TITLE OF INVENTION: Transcription Factors Related to TFIIA
;; FILE REFERENCE: 119941-1092
;; CURRENT APPLICATION NUMBER: US/10/156,634A
;; CURRENT FILING DATE: 2002-05-28
;; PRIOR APPLICATION NUMBER: 09/326,529
;; PRIOR FILING DATE: 1999-06-04
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 4

LENGTH: 1182
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-156-634A-4

Query Match 6.4%; Score 105.5; DB 15; Length 1182;
 Best Local Similarity 21.5%; Pred. No. 0.26;
 Matches 45; Conservative 43; Mismatches 84; Indels 37; Gaps 8;

19 FTQSPKSKVADLLGSFEKRR-----LLITPAKNNMYVQ-----RDYLSF 64
 268 FRSQKSGWSEFLRIPEKKNMSSSQWGPITFLKVLPGGILOMYEQLEKPEKFIQIDLPY 327
 65 CKMATRKISVITIEGPVNNSTKIDHFDLNKPKRVVDDVDLQRLISELRKEGMY 124
 328 CRLSPEKYENFSVAGKI--YKIEHVSYTEKR--KYSKTEVHNEPDIQMLKIGTSY 383
 125 NDFEVLTVDLRV-----KQYEVPTMKSVFDLIDTFQSRI---KDKKK 170
 384 HFDLDFLTVEELMKLPAVSKPKKNYEEOEISTL---EIVDNFNGKVTKEGFEVASVI 439
 171 EGIYCKEEVGVLELF--PINGSVVERED 198
 440 TOIYCLCFVNGNLGCEFLTLDLELPKDE 468

RESULT 10

US-09-883-797-6
 Sequence 6, Application US/09883797
 Patent No. US20020066123A1

GENERAL INFORMATION:
 APPLICANT: Jaworski, Jan G.
 APPLICANT: Post-Beltemmiller, Martha A.
 APPLICANT: Todd, James
 TITLE OF INVENTION: FATTY ACID ELONGASES
 FILE REFERENCE: 07148/064001
 CURRENT APPLICATION NUMBER: US/09/883,797
 CURRENT FILING DATE: 2001-06-18
 PRIOR APPLICATION NUMBER: 08/868,373
 PRIOR FILING DATE: 1997-06-03
 NUMBER OF SEQ ID NOS: 22
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 6
 LENGTH: 504
 TYPE: PRT
 ORGANISM: Arabidopsis thaliana
 US-09-883-797-6

Query Match 5.8%; Score 95.5; DB 9; Length 504;
 Best Local Similarity 20.3%; Pred. No. 0.72;
 Matches 68; Conservative 52; Mismatches 102; Indels 113; Gaps 16;

2 KGGKTEQDGYQKPTNKHFTQSPKSVADLLGSFEKRRLLITAKAENMYVOQREYL 61
 46 KGSKINVEDLQK--PSLHTQNNLQISLLFLVFWWILYMLTRPK-----PYLL 94
 62 ESF--CKM--ATTRKISVITIEGPVNNSTKIDHFDLNKPKRVVDDVDLQRLISELRK 118
 95 VDFSCYLPSPSHKVISIQRLMGHARRA-----R 121
 119 EYGMTYDNFENVYLDVLRVQYEVPTMKSVFDLIDTFQSRIKD---MEKOKKEGI 173
 122 EAGMKWKN-----KESDHLVD--FQKILERSGLGQETVIFPGL 158
 174 VCKEEVGVLELFPIN---GSSVVEREDVPAHLVADINNYQSEYFSMLYGVGDNV 230
 159 QC-----FPLQOGMGASRKETEIEFGALDNLRNRTGVKDDIGIIVV---NSS 204
 231 SWPSPMNSMNVIV--YDLIDSMQLRQEMAIQOOL---GMRQCKMSQAMVITVTKPTR 285
 205 TFNPPLSLASMTVKNYKLRDNI-----KSLNLGSGMS---AGVIAVDYAKGLL 250
 286 MYTRMTYIMRYITMDLTLEOKYVTLDSASFSC 320

Db 251 QVHRNTVAI--VVSPTENTQNLGLKKNKSMVLVTC 283

RESULT 11
 US-09-925-302-486
 Sequence 486, Application US/09925302
 Patent No. US20020044941A1

GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 FILE REFERENCE: PA104
 CURRENT APPLICATION NUMBER: US/09/925,302
 CURRENT FILING DATE: 2001-08-10
 PRIOR APPLICATION NUMBER: PCT/US00/05918
 PRIOR FILING DATE: 2000-03-08
 PRIOR APPLICATION NUMBER: 60/124,270
 PRIOR FILING DATE: 1999-03-12
 NUMBER OF SEQ ID NOS: 896
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 486
 LENGTH: 510
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (145)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-925-302-486

Query Match 5.8%; Score 95; DB 9; Length 510;
 Best Local Similarity 19.5%; Pred. No. 0.82;
 Matches 70; Conservative 63; Mismatches 126; Indels 100; Gaps 18;

26 SVADLLGSFEKRR--LLITPAKNNMYVOQREYLESCKMATRKISVITIEGPVNN 83
 120 SAGNITQLFHGKSRQRIINIINAKFXNLVFLKQ--VNTDFNIFIAVFGISTAMGMSL 178
 84 STMKIDHFDLNKPKRVVDDVDL-----ORLISEL--RKEGYMT---YN 125
 179 GIKGTEHQVHS-----LHFRDQVNASSKYEITTTIHMLFKLTHRLFRNFGYTLRSVN 233
 126 DF-----FMVLTVDLVRQYEVPTMKSVFDLID--TFQSRIKDKKKKEGIV--- 174
 234 DLYIQKPFILDFETKRYEY---FAEAQIADSDPAFISKTNNHIMKLKGLIKDALE 290
 175 -----CKEEVGVLELFPIN---NGSSVVEREDVPAHLVADINNYQVSP-- 215
 291 NIDPATQMMILNCITFKGSVNKKFPVEMTHNHNFRLENEREVYKVMQTKGNFLAANDOE 350
 216 -----EY---PSMLIV--GKQGNVKS-----WYDS-----PMMSM 240
 351 LDCDILQLEIVGGISMLIVPHKMSGKLTLEAQLTPRVERMKQSMTRNTRREVLLPKFL 410
 241 VYVVDLIDSMQLRQEMAIQOOL---LGRCKMSQAMV---TIVTKDTRMTV 293
 411 EKNYLVESLKLKMGIRMLFDPKNGNAGISDQRIALDLKHOGGITVNEGQATTVTV 469

RESULT 12
 US-09-815-242-5229
 Sequence 5229, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5229
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5229
```

```

Query Match          5.7%; Score 93.5; DB 9; Length 1184;
Best Local Similarity 22.7%; Pred. No.3.9;
Matches 59; Conservative 40; Mismatches 112; Indels 49; Gaps 11;
```

```

QY 7 EODGYQKPTNKHFTQSPKRSV-----ADLLGSEFGKRRLLLTAPKAENMY 53
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 309 ERKKQSEFTNARYEEQENLIELLENISNEISEADTYKSLSKQKELNAVIRELEQLY 368
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 54 V--QQRDEYLSFCGMATRKISVITIFGPVNNSTPKIDHFOLDNEKPRVVD-----E 105
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 369 VSDEHNDKLE---EIKNEYTYLMSQSDVNDIRFLKHTTEENAKKSRDLSLVEYFE 425
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 106 DLVVD-QRLISELRKEGYMTYNDFFWY-----LTDVDLRVKQYEVPTMKSVFDL 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 426 QLKDIQGIKTTKRYQQTNKELSAVDKEIKNIEKDLDTTKAONEYEE---KLQAVRY 482
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 IDTFOSRIKDEKQKKEIGVCEYGVY-----ELFPINGSSVVEREDVPAHLVKDIRN 209
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 483 TEKMTTRIDSLATQEEETFFNGVKHILKAKNKELGIGH-AVAEIIIDVSKLTQAIET 541
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 210 YFOVSEYFMSLVG--KDG 227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 542 ALGASLQH---VIYDSEKDG 558
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```

RESULT 13
US-09-815-242-12125
; Sequence 12125, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5229
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12125
```

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12125
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12125
```

```

Query Match          5.7%; Score 93.5; DB 9; Length 1188;
Best Local Similarity 22.7%; Pred. No.4;
Matches 59; Conservative 40; Mismatches 112; Indels 49; Gaps 11;
```

```

QY 7 EODGYQKPTNKHFTQSPKRSV-----ADLLGSEFGKRRLLLTAPKAENMY 53
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 310 ERKKQSEFTNARYEEQENLIELLENISNEISEADTYKSLSKQKELNAVIRELEQLY 369
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 54 V--QQRDEYLSFCGMATRKISVITIFGPVNNSTPKIDHFOLDNEKPRVVD-----E 105
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 370 VSDEHNDKLE---EIKNEYTYLMSQSDVNDIRFLKHTTEENAKKSRDLSLVEYFE 426
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 106 DLVVD-QRLISELRKEGYMTYNDFFWY-----LTDVDLRVKQYEVPTMKSVFDL 154
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 427 QLKDIQGIKTTKRYQQTNKELSAVDKEIKNIEKDLDTTKAONEYEE---KLQAVRY 483
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 155 IDTFOSRIKDEKQKKEIGVCEYGVY-----ELFPINGSSVVEREDVPAHLVKDIRN 209
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 484 TEKMTTRIDSLATQEEETFFNGVKHILKAKNKELGIGH-AVAEIIIDVSKLTQAIET 542
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 210 YFOVSEYFMSLVG--KDG 227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 543 ALGASLQH---VIYDSEKDG 559
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```

RESULT 14
US-09-815-242-11205
; Sequence 11205, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
```

```

1 PRIOR APPLICATION NUMBER: 60/269,308
2
3 PRIOR FILING DATE: 2001-07-16
4
5 NUMBER OF SEQ ID NOS: 14110
6
7 SOFTWARE: FASTSEQ for Windows Version 4.0
8
9 SEQ ID NO: 11205
10
11 LENGTH: 502
12
13 TYPE: PRY
14
15 ORGANISM: Haemophilus influenzae
16
17 OS-09-815-242-11205

```

Query Match	5.6%	Score 91.5	DB 9	Length 502
Best Local Similarity	21.4%	Pred. No. 1.8		
Matches 58, Conservative	40	Mismatches	92	Indels 81
				Gaps 9

[illegible]

```

RESULT 15
US-09-963-875-1
: Sequence 1, Application US/09963875
: Patent No. US20020164307A1
: GENERAL INFORMATION:
: APPLICANT: Massachusetts General Hospital
: TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating
: TITLE OF INVENTION: Mellitus
: FILE REFERENCE: 17633/1235
: CURRENT APPLICATION NUMBER: US/09/963,875
: CURRENT FILING DATE: 2001-09-26
: PRIOR APPLICATION NUMBER: US60/169082
: PRIOR FILING DATE: 1999-12-06
: PRIOR APPLICATION NUMBER: US 60/215109
: PRIOR FILING DATE: 2000-06-28
: PRIOR APPLICATION NUMBER: US 60/238880
: PRIOR FILING DATE: 2000-10-06
: PRIOR APPLICATION NUMBER: US 09/731261
: PRIOR FILING DATE: 2000-12-06
: NUMBER OF SEQ ID NOS: 58
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 1618
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-963-875-1

```

Query Match	5.4%	Score	89.5	DB	10	Length	1618
Best Local Similarity	25.3%	Pred. No.	15				
Matches	41	Conservative	26	Mismatches	80	Indels	15
						Gaps	6

```

0Y      13 KPTNHFT---OSPKSVADIGSEFGKRRLLITAPKNNMNYVQOQREYSEFCMKAT 69
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      614 KPTGKEDQYTLQSLCKENGEIMKSLBGNLFTLF--PCHENGELVSLQENLESL--TAL 669
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
0Y      70 KRISITITFGPNNSTMTKIDHQLDNEKMRVVDDEDIYDQRLISLREKGYMTYDFEM 129
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      670 EKENOEPLSPVGDGEALRPLTKENOEPLRSLD---NKEAFSLSE---NOEPLK 722
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

```
QY      130 VLTVDVLRVKQYEEVPITMKSVDLIDTFQSRIKOMEKQKE 171
      | : | : | 11 : | : : 11 : :
Db      723 TLEEDQSIVRPLETE-NHKSLSRLSEQDQETLRTLLEKEFQQ 763
```

Search completed: July 24, 2003, 13:14:52
Job time : 23.6724 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:48:39 ; Search time 14.1534 Seconds
(without alignments)
956.622 Million cell updates/sec

Title: US-09-884-319A-4
Perfect score: 1645
Sequence: 1 KKGKTEQDGYQKPTNKHFT.....DLTRQKYVLDSASFICSC 320

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1645	100.0	320	1	US-08-726-525-4	Sequence 4, Appl1
2	1645	100.0	320	2	US-08-487-942-4	Sequence 4, Appl1
3	1645	100.0	320	2	US-08-726-036A-4	Sequence 4, Appl1
4	1645	100.0	320	4	US-09-083-516-4	Sequence 4, Appl1
5	396.5	24.1	358	1	US-08-253-155A-36	Sequence 36, Appl1
6	105.5	6.4	1182	4	US-09-326-529-4	Sequence 4, Appl1
7	95.5	5.8	504	4	US-08-868-373-6	Sequence 6, Appl1
8	94	5.7	505	3	US-09-257-581-5	Sequence 5, Appl1
9	94	5.7	505	3	US-09-257-581-7	Sequence 7, Appl1
10	89.5	5.4	514	4	US-09-107-532A-6282	Sequence 6282, Ap
11	89.5	5.4	1211	4	US-09-134-001C-4820	Sequence 4820, Ap
12	89.5	5.4	1618	1	US-07-853-913-4	Sequence 4, Appl1
13	89.5	5.4	1657	1	US-08-287-959-1	Sequence 1, Appl1
14	89	5.4	644	4	US-09-198-452A-63	Sequence 63, Appl1
15	89	5.4	985	4	US-08-680-326-41	Sequence 41, Appl1
16	88.5	5.4	618	3	US-09-299-378-4	Sequence 4, Appl1
17	88.5	5.4	1145	4	US-09-470-443-2	Sequence 2, Appl1
18	88.5	5.4	1145	4	US-09-470-443-4	Sequence 4, Appl1
19	86	5.2	414	2	US-08-845-161A-2	Sequence 2, Appl1
20	86	5.2	414	3	US-09-270-751-2	Sequence 2, Appl1
21	86	5.2	414	4	US-09-168-218B-4	Sequence 4, Appl1
22	85.5	5.2	502	4	US-09-328-352-5891	Sequence 5891, Ap
23	85.5	5.2	2710	1	US-08-480-604A-6	Sequence 6, Appl1
24	85.5	5.2	2710	2	US-08-405-496A-6	Sequence 6, Appl1
25	85.5	5.2	2710	3	US-08-915-136-6	Sequence 6, Appl1
26	85.5	5.2	2710	4	US-08-957-310-6	Sequence 6, Appl1
27	85.5	5.2	2710	4	US-10-011-366-6	Sequence 6, Appl1

28	85	5.2	793	2	US-08-468-558-5	Sequence 5, Appl1
29	85	5.2	793	3	US-08-676-444-5	Sequence 5, Appl1
30	85	5.2	1584	3	US-09-251-645-6	Sequence 6, Appl1
31	84.5	5.1	488	4	US-09-134-001C-4246	Sequence 4246, Ap
32	83.5	5.1	467	2	US-08-686-599A-17	Sequence 17, Appl1
33	83	5.0	378	4	US-09-107-532A-3777	Sequence 3777, Ap
34	83	5.0	790	4	US-09-634-957-2	Sequence 2, Appl1
35	83	5.0	790	4	US-09-723-153-2	Sequence 2, Appl1
36	83	5.0	790	4	US-09-723-429-2	Sequence 2, Appl1
37	82.5	5.0	516	4	US-09-107-532A-7156	Sequence 7156, Ap
38	82.5	5.0	696	4	US-09-107-532A-5205	Sequence 5205, Ap
39	82	5.0	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
40	82	5.0	1388	2	US-08-685-576-1	Sequence 1, Appl1
41	81.5	5.0	482	2	US-08-686-599A-18	Sequence 18, Appl1
42	81.5	5.0	460	4	US-09-252-991A-22255	Sequence 22255, A
43	81.5	5.0	493	2	US-08-686-599A-5	Sequence 5, Appl1
44	81.5	5.0	493	2	US-08-686-599A-16	Sequence 16, Appl1
45	81.5	5.0	501	1	US-08-451-715A-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1
US-08-726-525-4
; Sequence 4, Application US/08726525
; Patent No. 5789181
GENERAL INFORMATION:
; APPLICANT: Lin, Lin-Ling
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; NUMBER OF INVENTION: BINDING
; CORRESPONDENCE ADDRESS: 7
; ADDRESS: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,525
; FILING DATE: 07-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,942
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-726-525-4
Query Match 100.0%; Score 1645; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.1e-177;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKGKTEQDGYQKPTNKHFTQSPKKSVAIDLGSFEGKRRLLITAPKAENNMVVOORDEX 60

Db	1	KKGGKTQDQGYÖKRPNTKHHFTQSPKKSVADLLSGFECKRLLLTATKAENMMVVOQRDEX	60
QY	61	LESFCKMATEKRIISVITIEFPVNNSTYKIDHFQDNEKPMKRVUDEDLVDORLISELRKEY	120
Db	61	LESFCKMATEKRIISVITIEFPVNNSTYKIDHFQDNEKPMKRVUDEDLVDORLISELRKEY	120
QY	121	GMRYNDEPMVLTVDLDAVKQYEVPTMKSVFLLIDTFOSRIKDMERÖKKEGIVCKEEYG	180
Db	121	GMRYNDEPMVLTVDLDAVKQYEVPTMKSVFLLIDTFOSRIKDMERÖKKEGIVCKEEYG	180
QY	181	GVLELFPINSSVYEREDVPAHLVKDIRNFQVSPXYFSMLVKGKGNYSWTPSPMMSM	240
Db	181	GVLELFPINSSVYEREDVPAHLVKDIRNFQVSPXYFSMLVKGKGNYSWTPSPMMSM	240
QY	241	VIVYDLDISQMLRQENAIQOOSLGMRCKMSQMAWTVITTKDJRTAVTMTVIMRVITM	300
Db	241	VIVYDLDISQMLRQENAIQOOSLGMRCKMSQMAWTVITTKDJRTAVTMTVIMRVITM	300
QY	301	DTLFEÖKRYVTLDSASFLLCSC 320	
Db	301	DTLFEÖKRYVTLDSASFLLCSC 320	

RESULT 2
 US-08-487-942-4
 Sequence 4, Application US/08487942
 Patent No. 5817476
 GENERAL INFORMATION:
 APPLICANT: Lin, Lih-Ling
 APPLICANT: Graham, James
 TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
 TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
 TITLE OF INVENTION: BINDING
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: MA
 COUNTRY: USA
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,942
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Brown, Scott A.
 REGISTRATION NUMBER: 32,724
 REFERENCE/DOCKET NUMBER: G15258
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 498-8224
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 320 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-487-942-4

Query Match	100.0%;	Score 1645;	DB 2;	Length 320;
Best Local Similarity	100.0%;	Pred. No. 1.1e-177;		
Matches 320;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 KKGATGDDGQKPTNKHFTQSPKKSVAADLLGFEGRRLLLTPAPKAENMMYVQORDEY 60
 |||||
 Db 1 KKGATGDDGQKPTNKHFTQSPKKSVAADLLGFEGRRLLLTPAPKAENMMYVQORDEY 60

QY	6	LESECKMAARKISVITIEGEPVANNMUKIDHOLONEKPMRVYDEDDIVDORLISELRKEY	120
Db	61	LESECKMAARKISVITIEGEPVANNMUKIDHOLONEKPMRVYDEDDIVDORLISELRKEY	120
QY	121	GMTYNDPFMYLIDVDLVRKYQYVEVITMKSVFDLIDTFOSRIKDMERKKEGIVCKEYVG	180
Db	121	GMTYNDPFMYLIDVDLVRKYQYVEVITMKSVFDLIDTFOSRIKDMERKKEGIVCKEYVG	180
QY	181	GVLELPIGSSVVEEEDVPAHLVMDINRYOVSPESYMLVKKDGNVKSMTYPSPMWSM	240
Db	181	GVLELPIGSSVVEEEDVPAHLVMDINRYOVSPESYMLVKKDGNVKSMTYPSPMWSM	240
QY	241	VIVYDIDLSMQLRQEMAIQOOSLGRCOKMSMOAVYITVTKDTRMYT RMTYVIMRYITM	300
Db	241	VIVYDIDLSMQLRQEMAIQOOSLGRCOKMSMOAVYITVTKDTRMYT RMTYVIMRYITM	300
QY	301	DTLEOKYVTLDSASFJCSG	320
Db	301	DTLEOKYVTLDSASFJCSG	320

```

1      RESULT 3
2      US-08-726-036A-4
3      Sequence 4, Application US/08726036A
4      Patent No. 5981482
5      GENERAL INFORMATION:
6      APPLICANT: Lin, Lih-Ling
7      APPLICANT: Graham, James
8      TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
9      TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
10     TITLE OF INVENTION: BINDING
11     NUMBER OF SEQUENCES: 7
12     CORRESPONDENCE ADDRESS:
13     ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
14     STREET: 87 Cambridgepark Drive
15     CITY: Cambridge
16     STATE: MA
17     COUNTRY: USA
18     ZIP: 02140
19     COMPUTER READABLE FORM:
20     MEDIUM TYPE: Floppy disk
21     COMPUTER: IBM PC compatible
22     OPERATING SYSTEM: PC-DOS/MS-DOS
23     SOFTWARE: PatentIn Release #1.0, Version #1.25
24     CURRENT APPLICATION DATA:
25     APPLICATION NUMBER: US/08/726,036A
26     FILING DATE:
27     CLASSIFICATION: 435
28     ATTORNEY/AGENT INFORMATION:
29     NAME: Brown, Scott A.
30     REGISTRATION NUMBER: 32,724
31     REFERENCE/DOCKET NUMBER: G15258
32     TELECOMMUNICATION INFORMATION:
33     TELEPHONE: (617) 498-8224
34     TELEFAX: (617) 876-5851
35     INFORMATION FOR SEQ ID NO: 4:
36     SEQUENCE CHARACTERISTICS:
37     LENGTH: 320 amino acids
38     TYPE: amino acid
39     TOPOLOGY: linear
40     MOLECULE TYPE: protein
41     US-08-726-036A-4

```

Query Match	100.0%;	Score 1645;	DB 2;	Length 320;
Best Local Similarity	100.0%;	Pred. No. 1.1e-177;		
Matches 320;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

1 KKGKTEQDGYQKPTNKHFTQSPKKSADLLGSFEGKRLLITAPKAENNMVYVQQRDEY 60

Db 1 KKGKTEQDGYQKPTNKHFTQSPKKSADLLGSEFGKRRLLLITAPAEENMYVQQRDEY 60

Qy 61 LESFCMATRKISVITIEGPNVNSTMKIDHFQLDNEKPMRVVDEDEDLVDRRLISELRKEY 1200

Db 61 LESFCMKATRKISVITITFGPVNNSMTKIDHFOLDNEKPMRVVDEDEDLVDRLLSELREKEY 120
QY 121 GMTYNDFEYVLTVDLVRKQYVEVPITMKSVFDLIDFOSRIKMEKQKEGIVCKEEVG 180
Db 121 GMTYNDFEYVLTVDLVRKQYVEVPITMKSVFDLIDFOSRIKMEKQKEGIVCKEEVG 180
QY 181 GVLELFPINGSVVEREDVPAHLVKDIRNFQVSPFEFSMLLVGKDGNAVSWTSPPMWSM 240
Db 181 GVLELFPINGSVVEREDVPAHLVKDIRNFQVSPFEFSMLLVGKDGNAVSWTSPPMWSM 240
QY 241 VIYYDLIDSQMLRQEKAIQOSLGMRCQKMSQAMVTITVTKDTRMTVIMRYITM 300
Db 241 VIYYDLIDSQMLRQEKAIQOSLGMRCQKMSQAMVTITVTKDTRMTVIMRYITM 300
QY 301 DTLTEOKRYVTLDLSASFCLSC 320
Db 301 DTLTEOKRYVTLDLSASFCLSC 320

RESULT 4

US-09-083-516-4
; Sequence 4, Application US/09083516
; Patent No. 6300086
; GENERAL INFORMATION:
; APPLICANT: Lin, Lin-Ling
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,516
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,942
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-083-516-4

Query Match

Best Local Similarity 100.0%; Score 1645; DB 4; Length 320;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGGKTEBODGYOKPTNKHFTQSPKSVADLIGSPEGRRLTLITAPKAENNMVVOQDDEY 60
Db 1 KKGGKTEBODGYOKPTNKHFTQSPKSVADLIGSPEGRRLTLITAPKAENNMVVOQDDEY 60

QY 61 LESFCMKATRKISVITITFGPVNNSMTKIDHFOLDNEKPMRVVDEDEDLVDRLLSELREKEY 120
Db 61 LESFCMKATRKISVITITFGPVNNSMTKIDHFOLDNEKPMRVVDEDEDLVDRLLSELREKEY 120
QY 121 GMTYNDFEYVLTVDLVRKQYVEVPITMKSVFDLIDFOSRIKMEKQKEGIVCKEEVG 180
Db 121 GMTYNDFEYVLTVDLVRKQYVEVPITMKSVFDLIDFOSRIKMEKQKEGIVCKEEVG 180
QY 181 GVLELFPINGSVVEREDVPAHLVKDIRNFQVSPFEFSMLLVGKDGNAVSWTSPPMWSM 240
Db 181 GVLELFPINGSVVEREDVPAHLVKDIRNFQVSPFEFSMLLVGKDGNAVSWTSPPMWSM 240
QY 241 VIYYDLIDSQMLRQEKAIQOSLGMRCQKMSQAMVTITVTKDTRMTVIMRYITM 300
Db 241 VIYYDLIDSQMLRQEKAIQOSLGMRCQKMSQAMVTITVTKDTRMTVIMRYITM 300
QY 301 DTLTEOKRYVTLDLSASFCLSC 320
Db 301 DTLTEOKRYVTLDLSASFCLSC 320

RESULT 5

US-08-253-155A-36
; Sequence 36, Application US/08253155A
; Patent No. 5691147
; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jenio
; TITLE OF INVENTION: CDK4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253,155A
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-253-155A-36

Query Match

Best Local Similarity 24.1%; Score 396.5; DB 1; Length 358;
Matches 86; Conservative 9; Mismatches 22; Indels 19; Gaps 3;

QY 55 QQRDEYLESEFCMKATRKISVITITFGPVNNSMTKIDHFOLDNEKPMRVVDEDEDLVDRLLS 114
Db 6 QQRDEYLESEFCMKATRKISVITITFGPVNNSMTKIDHFOLDNEKPMRVVDEDEDLVDRLLS 54
QY 115 ELRKEVGMATND---FFMYVLTVDLVRKQYVEVPITMKSVFDLIDFOSRIKMEKQK 170
Db 55 ELRKEVGMATND---FFMYVLTVDLVRKQYVEVPITMKSVFDLIDFOSRIKMEKQK 110

QY 171 EGIVCKEEVGVLELF 186
Db 111 EGIVCKEDKKQSLLENF 126

RESULT 6
US-09-326-529-4
Sequence 4, Application US/09326529
Patent No. 6448072
GENERAL INFORMATION:
APPLICANT: Delong Dr., Jeff
TITLE OF INVENTION: Transcription Factors Related to TFIIA
FILE REFERENCE: 119941-1053
CURRENT APPLICATION NUMBER: US/09/326,529
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1182
TYPE: PRT
ORGANISM: Homo sapiens
US-09-326-529-4

Query Match 6.4%; Score 105.5; DB 4; Length 1182;
Best Local Similarity 21.5%; Pred. No. 0.02;
Matches 45; Conservative 43; Mismatches 84; Indels 37; Gaps 8;

QY 19 FQSPKSVADLLGSEFKRR-----LLLTAPKAENMYVQ-----RDEVLESP 64
Db 268 FRSQPSGSGSEFMRIPEKKNMSSROMGPIFLVLPGLIOMIYEGLEKPFKEIQIDPY 327
QY 65 CKMATRKISVITTFGVNNSTMTKIDHFDLNEKPMRVVDEDDIVDRLISELRKREYGMTY 124
Db 328 CRLEPKVEKVESVAKGIH--TVKIEHVSYTEKR--KYHSTEVVHEPDIDQMLKLGSTSY 383
QY 125 NQFPMVLTVDLRY-----KOYEPVITMKSVFDLIDTFQSKRI---KDMEKOKR 170
Db 384 HPELDLFTVVEEELMKLPAVSKRPKNYERQEISL---ELVDNFMGKVKYKGEKVESAVI 439
QY 171 EGIVCKEEVGVLELF-PINGSSVERED 198
Db 440 TQIYCLCFVNGNLECFITLNDLELPKRDE 468

RESULT 7
US-08-868-373-6
Sequence 6, Application US/08868373
Patent No. 6307128
GENERAL INFORMATION:
APPLICANT: Jaworski, Jan G.
APPLICANT: Post-Beltemiller, Martha A.
APPLICANT: Todd, James
TITLE OF INVENTION: FATTY ACID ELONGASES
FILE REFERENCE: 07148/064001
CURRENT APPLICATION NUMBER: US/08/868,373
CURRENT FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 504
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-08-868-373-6

Query Match 5.8%; Score 95.5; DB 4; Length 504;
Best Local Similarity 20.3%; Pred. No. 0.068;
Matches 68; Conservative 52; Mismatches 102; Indels 113; Gaps 16;
QY 2 KGGKTQDQYQKPTNHFQSPKSVADLLGSEFKRRLLLTAPKAENMYVQORDEYL 61
Db 46 KGSKIVEDLQK-FSLHHTQNNLQITISLLFLVVFWMILMLTRPK-----PVYL 94

QY 62 ESF-CKM--ATRKISVITTFGVNNSTMTKIDHFDLNEKPMRVVDEDDIVDRLISELRK 118
Db 95 VDESCYLPSPSHLKVSIOTLMGHARRA-----R 121
QY 119 EYGMTYNDEFFMVLTDVLDKQYVEPITMKSVFDLIDTFQSKRI---MEKOKKEGI 173
Db 122 EAGMCWKN-----KESDHLVD--FOEKILERSGLQGETYIPEGL 158
QY 174 VCKEEVGVLELFPIN---GSSVEREDYPAHLVDIRVYFOVSPFMSMLVKGDNVAK 230
Db 159 QC-----FPLQCGKASRKREVEYIFGALDLPFRNTGVKPPDIGILVY---NRS 204
QY 231 SWYSPSMNVIV--YDLIDSMQLRQEMALQOST---GMRCQKMSQAMVTITVTKDTR 285
Db 205 TFPPTSLASMIYKYLKLDNI-----KSLMIGMGCS---AGVAVDVAKGLL 250
QY 286 MYTRMTVIMRVITMDLTREQKYVTLDSASFLQSC 320
Db 251 QVHRNTYAI--VVSTENITQNLTKGNKSMALVTNC 283

RESULT 8
US-09-257-581-5
Sequence 5, Application US/09257581
Patent No. 6207419
GENERAL INFORMATION:
APPLICANT: Church, Frank C.
APPLICANT: Bauman, Susannah J.
TITLE OF INVENTION: THROMBIN INHIBITORY AGENTS AND METHODS OF USING SAME
FILE REFERENCE: 5470-232
CURRENT APPLICATION NUMBER: US/09/257,581
CURRENT FILING DATE: 1999-02-25
EARLIER APPLICATION NUMBER: 60/076,210
EARLIER FILING DATE: 1998-02-27
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 505
TYPE: PRT
ORGANISM: Homo sapiens
US-09-257-581-5

Query Match 5.7%; Score 94; DB 3; Length 505;
Best Local Similarity 19.5%; Pred. No. 0.1;
Matches 70; Conservative 63; Mismatches 126; Indels 100; Gaps 18;

QY 26 SVADLLGSEFKRRLL--LLLTAPKAENMYVQORDEYLSFCKMATRKISVITTFGVNN 83
Db 109 SAGNITQLFHKRSRIQRLINLAKKFAFLRYLAKD--VTFDNIFTAPYGISTAMGMSL 167
QY 84 STMKIDHFDLNEKPMRVVDEDDIVD-----QRLISEL-RKEYGMT--YN 125
Db 168 GKJGTFHEQVHS-----LHFKDFVNASSKEYEITTHNLFRKLTGHLFPRNGCYLTRSVN 222
QY 126 DF-----FNVLDVLDLRYKQYVEPITMKSVFDLID-TQSRKDKMEKOKKEGIV----- 174
Db 223 DLYIQKQFPLIDFERKRVREY---FAEAQIADFSQPARISKNNHIMLTGLKINDALE 279
QY 175 -----CKEEVGVLELFPIN---GSSVEREDYPAHLVDIRVYFOVSP- 215
Db 280 NIDPATQMLINLTIRKGSVKNKFPVEMTHNHNFRINEREVYKVSMTQKGNFLAANDOE 339
QY 216 -----EY--FSMLIV--GKDGNVKS-----WYPS-----PMWGM 240
Db 340 LODCDILQLEVYGSISMLIIVPHKMSGKLTLEAQLTPRYVERMQSKMNTREVILPKFLL 399
QY 241 VIYVDLIDSMQLRQEMALQOST---LGMRCQKMSQAMV---TIVTKTFRMYTRMTTY 293
Db 400 EKNYNLVESIKLGMIRLFDKNGNMAGISDORIAIDLFRHQGITVNEBGTQATVTVTV 458
RESULT 9
US-09-257-581-7


```

? Sequence 7 Application US/09257581
? Patent No. 6207419
? GENERAL INFORMATION:
? APPLICANT: Church, Frank C.
? APPLICANT: Bauman, Susannah J.
? TITLE OF INVENTION: THROMBIN INHIBITORY AGENTS AND METHODS OF USING SAME
? FILE REFERENCE: 5470-233
? CURRENT APPLICATION NUMBER: US/09/257,581
? CURRENT FILING DATE: 1999-02-25
? EARLIER APPLICATION NUMBER: 60/076,210
? EARLIER FILING DATE: 1998-02-27
? NUMBER OF SEQ. ID NOS.: 7
? SOFTWARE: PatentIn Ver. 2.0
? SEQ. ID NO. 7
? LENGTH: 505
? TYPE: PR1
? ORGANISM: Homo sapiens
? US-09-257-581-7

```

Query Match	5.7%;	Score 94;	DB 3;	Length 505;
Best Local Similarity	19.5%;	Pred. No. 0.1;		
Matches 70;	Conservative 63;	Mismatches 126;	Indels 100;	Gaps 18

```

QY 26 SVADLLIGSEGGRRLL--LLITAPKKNMMYUOOREYUESCKAKTRISVITTEFGVNN 83
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 109 SAGNIILQLEPHGSRIGRLNLIILAKAFNLIYRLKQ--VTFEDNFIAPVGSTAMGMSL 167
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 84 STYKIDHFOIDNEKPRRVVDEDDLYD-----ORLSEL--RKEYGMT---YN 125
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 168 GLKGETHEGVHS-----LIHFQDFNASSKEYEITTHNLFPKRLIHLRFRFRNFQITLRSVN 222
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 126 DF-----FVILNDVDRVKKQYEVYITMKSVFDLID--TFQSRIKDMCKKEGV----- 174
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 223 DLYIQKQFILLDFRKKVREYV---FAEQIADFSDFAFISKTNNHIMLKIGLKNAL 279
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 175 -----CKEEVGVLELFPY-----NGSSVEREDVPAHLKYKDRIYQVSP-- 215
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 280 NIDPATOMILNCYIFFGSGWVKPFVEPETHIHNFRLNEREVVKKVSMOTKGNFLAANDOE 339
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 216 -----EY-----FSMILV--GKDDNVVS-----WYPS-----PMWSM 240
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 340 LDCDILQLEFVGGISMLIVVPHKMSGMTLEAQLPRVYERVMQSMNRRREVLLLPFKL 399
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 241 VIYYIDLIDSMQLRQDMAIOS---LGNRCQKMSQAVN---TIYTKDTRMYTRMYTV 293
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 400 EKNYNILVESIKLGMIGIEMLEFDKGNMAGISDRIADLEFKHOGITTVNEEGQATVTVTV 458
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 10
 US-09-107-532A-6282
 Sequence 6282, Application US/09107532A
 Patent No. 6583275
 GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107,532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:

```

1      APPLICATION NUMBER: 60/085,598
2      FILING DATE: 14 May 1998
3      APPLICATION NUMBER: 60/051571
4      FILING DATE: July 2, 1997
5      ATTORNEY/AGENT INFORMATION:
6      NAME: Atinello, Pamela Deneke
7      REGISTRATION NUMBER: 40,489
8      REFERENCE/DOCKET NUMBER: GTC-012
9      TELECOMMUNICATION INFORMATION:
10     TELEPHONE: (781)893-5007
11     TELEFAX: (781)893-8277
12     INFORMATION FOR SEQ ID NO: 6282:
13     SEQUENCE CHARACTERISTICS:
14     LENGTH: 514 amino acids
15     TYPE: amino acid
16     TOPOLOGY: linear
17     MOLECULE TYPE: protein
18     HYPOTHEetical: YES
19     ORIGINAL SOURCE:
20     ORGANISM: Enterococcus faecium
21     FEATURE:
22     NAME/KEY: misc_feature
23     LOCATION: (B) LOCATION 1...514
24     SEQUENCE DESCRIPTION: SEQ ID NO: 6282:
US-09-107-532A-6282

```

Query Match	5.48;	Score 89.5;	DB 4;	Length 514;
Best Local Similarity	20.08;	Pred. No. 0.34;		
Matches 52;	Conservative 52;	Mismatches 87;	Indels 69;	Gaps 12

```

QY      25 KSVADILGSEFGKRRLLLTATPAKANNMYVOODIEYSEFCMKATRKISVITIGEPVNN 84
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db      230 RETADLIDS-----LYDLEREMEAQRVKNQSTIAEYIKHTK-----NNR 272

QY      85 TM--KIDH----FOLDNEKPMRY-----VDEDELVDQRLLSELRK-----EYGMTYND 126
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db      273 QLLIEDITQSAYTLNHNIEGRVGRFQTEVEEMERQNMQMPQJROHEIPYSEVRFPYKT 332

QY      127 FFWLIDVDLIRKQYEVPI TM-----KSVFDLIDFPQSRIDM-----EKQKEGIVC 175
      | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      333 VFVVLDEIE--KQQVEIDDSLHELKRGKEQENIDTFEERKLSLRFVEKQMLPBL-- 387

QY      176 KEEVGVALFPI NGSSVVEREDVPAHLVKDIRNFQVSP EYF5MLLVGDKGNYSWPS 235
      | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      388 ---PNDYLEFFEFVATDRIE-----LSVLNKRIVNMEEVRIALACEEDLELDKK 436

QY      236 PMMSMVIYVDLIDSMQLRRQ 255
      : | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db      437 -----THDLVDAALTEQ 449

```

RESULT 11
 US-09-134-001C-4820
 : Sequence 4820, Application US/09134001C
 : Patent No. 6380370
 : GENERAL INFORMATION:
 : APPLICANT: Lynn Doucette-Stamm et al
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 : TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: GFC-007
 : CURRENT APPLICATION NUMBER: US/09/134,001C
 : PRIOR FILING DATE: 1998-08-13
 : PRIOR APPLICATION NUMBER: US 60/064,964
 : PRIOR FILING DATE: 1997-11-08
 : PRIOR APPLICATION NUMBER: US 60/055,779
 : PRIOR FILING DATE: 1997-08-14
 : NUMBER OF SEQ ID NOS: 5674
 : SEQ ID NO 4820
 : LENGTH: 1211
 : TYPE: prt
 : ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4820

Query Match	Similarity	52.4%	Score 89.5	DB 4	Length 1211
Best Local	Similarity	22.3%	Score No.1.4		
Matches	Conservative	45	Mismatches	91	Indels 77; Gaps 13
QY	7	EODGYQAKTNNKHFQD-----	SPKKSADVLGSEGGKRRLLITAPAKENNNY	53	
DB	332	ERKNQOSTNNARYEKEIDNLESQIDSIKNEKAQNEKLADLNKKOKQINKEVEQLESLLY	391		
QY	54	V-QORDEYLSFCMKATRKISVLTIFGP-----	VNNSTWKIDHFOLDNE-KPMRYVDDIDL	107	
DB	392	ISDQHDKELEI-----	KNSYVTLMSQSVNNNDIFLEHTIENENAKKSRL--DSRL	443	
QY	108	VD-----	QRLSELKKEYGMT-----	YNDFEMLTIDVDLRVQOY	142
DB	444	VEAFNQDKDIOONTOTOTQKEYOSSKSKSEKVEQNIQOLEQOLTDSKRLSEENKLYAY	503		
QY	143	EVPIYTKMSVFLPIDP-----	FOSRIMDEKOKKEGIVCKEEVGVLELFPINGSAYE	195	
DB	504	RYNEKLKGRISDLATQEDDYTYFFNGVAKHLLAKOK-----	ELGIGH-AVAE	550	
QY	196	REDVPAHLVKDIRNFQVSPREYFSLIAG--KDG	227		
DB	551	VINVPSEMTQAIETALGASIQH--	VIQDNEGD	581	

```

;          TOPOLOGY: linear
;          MOLECULE TYPE: protein
US-07-853-913-4

Query Match          5.4%; Score 89.5; DB 1; Length 1618;
Best Local Similarity 25.3%; Pred. No. 2.2;
Matches 41; Conservative 26; Mismatches 80; Indels 15; Gaps 6;

Oy      13 KPTNKHFP---QSPKKSVAADLLGSEPKRRLLLTAPAEENMMVYQQRDEYLESCKMAT 69
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      614 KPTGKEDTQTLQSLCKENQELMKSLGKULFFLF--PTEQNELVSSLQEMNESL--TAL 669

Oy      70 KISVITIEGVNNSTMKIDHFOLDNEKPMRVYDDEDLVDQRLISELRKEXGMTYNDFEM 129
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      670 EKENQEPKRSPEVGDEALRPLTKENQEPKLSLEDE---NNEAFRSLEKE----NOEPLK 722

Oy      130 VLTVDLRLVKQYVEPIFMKSVFPLDIFQSRIDKMEOKKE 171
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      723 TLEEDQSIVRPLETE--NHKSLRSLSEDDQETTLRLKEEQO 763

```

[illegible]

```

Db      749 ITRLOARCGYLROEFBSRNFELKKOIPAITCISQMRGKQKKAQDRLAIRSHKDE 808
QY      74 VTII-----FGPVNSTMKIDHFOJDN-----EKPMRV 101
Db      809 VVKIOSLAMHQAARKRYRDLQYFRDHIDIIKIOFIAANKARDDYKTLINAEDEPMV 868
QY      102 V-----DDEDVQDLISELRKEY-----GMTYDFEMLTVDLHVQKQYEV 145
Db      869 VKRFVHLLDQSPDPOEELDMKKMREEVTLIRSNQOENDLINDIKIGLAVK-----K 924
QY      146 TTKSVFEDLIDTFOSIKMEKQKEGI-----VCKEEVGVLLEFPINGSVVEREDVA 201
Db      925 IYLODVY-----SHSKLTFTKKNKQOLSDMMINKQKGLKAL-----SKEKREKLEA 971
QY      202 HLYKDIRNFQVSPFESMLLVGKQGNVKSMTSPKMSVVIYDILDSMQLRQEMAIQO 261
Db      972 Y--QHLEFYLLQNTPIYLAFLIFQMPON-----KSTKEMDSVLEFTLYNVA5NOREYTL-- 1022
QY      262 SIGMRCQKSMQAMVTITVTKDTRMTVIMRYTMD 301
Db      1023 ---LRFKTALEIKSKYDQIQEIVTGNPTVYKMYVSN 1059

```

```

RESULT 14
US-09-198-452A-63
: Sequence 63, Application US/09198452A
: Patent No. 6559294
: GENERAL INFORMATION:
: APPLICANT: Griffiths, R.
: TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
: TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
: FILE REFERENCE: 9710-003-999
: CURRENT APPLICATION NUMBER: US/09/198,452A
: CURRENT FILING DATE: 1998-11-24
: NUMBER OF SEQ ID NOS: 6849
: SEQ ID NO 63
: LENGTH: 644
: TYPE: PRN
: ORGANISM: Chlamydia pneumoniae
: US-09-198-452A-63

```

```

Query Match      5.4%; Score 89; DB 4; Length 644;
Best Local Similarity 21.5%; Pred. No. 0.55;
Matches 55; Conservative 36; Mismatches 95; Indels 70; Gaps 9;

```

```

QY      49 ENNMVQODEYLESCKMATRKISVITJFGPVNSTMKI-----DHFOLDNEKPMR 100
Db      279 ETCLVKSILTKSYRDFCDYKRAKIL-----PDENNSARAEOFRREVKDHMEDLANETVFM 332
QY      101 VYDD-----EDLVQDLISELRKEYGMTYNDFFVYLDVLDLVKQYEVPI 146
Db      333 VVEDGRIDIEVLTVANGMPDRYPEHLILERKDKVSHQ-----LWEATIRVKE-ALVTV 386
QY      147 TTKSVFEDLIDTFOSIKMEKQKEGI-----KEEVGVLELFPINGSV 194
Db      387 SVARVAFERKDGQGNQKQKQEKTERLRLKDLRQEGHRAQERLEKITALVYEVSVV 446
QY      195 EREDVPAHLVKDIRNFQVSPFESMLLVGKQGNVKSMTSPKMSVVIYDILDSMQLRR 254
Db      447 ETERR-----ERKFNLEKAY-----GNLEERYQS-----VYDQEDDYWTJEQK 482
QY      255 QEMAIQOSLGMRCQKM 270
Db      483 NREAEFRKAGTKYVRSK 498

```

```

RESULT 15
US-08-680-326-41
: Sequence 41, Application US/08680326
: Patent No. 5925733
: GENERAL INFORMATION:
: APPLICANT: ROSE, TIMOTHY M.

```

```

: APPLICANT: BOSCH, MARINX
: APPLICANT: STRAND, KURT
: APPLICANT: TODARO, GEORGE J.
: TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
: TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
: NUMBER OF SEQUENCES: 152
: CORRESPONDENCE ADDRESS:
: ADDRESSER: MORRISON & FOERSTER
: STREET: 755 Page Mill Road
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94304-1018
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/680,326
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Schiff, J. Michael
: REGISTRATION NUMBER: 40,253
: REFERENCE/DOCKET NUMBER: 29938-20001.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 813-5600
: TELEFAX: (415) 494-0792
: TELEX: 706141
: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 985 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-680-326-41

```

```

Query Match      5.4%; Score 89; DB 2; Length 985;
Best Local Similarity 20.5%; Pred. No. 1.1;
Matches 40; Conservative 38; Mismatches 75; Indels 42; Gaps 7;

```

```

QY      96 EKPMRVDDDEDVQDLISELRKEYGMTYNDFFVYLDVLDLVKQYEVPIITKSVFEDLI 155
Db      298 ENPMEMIPVPHL-----HLDASRIKFEACKNERFMILAFIN-RLKSVNVLVIYNAQFD-I 351
QY      156 DTFOSRIKDMKQKKEGIYCKEEVGVLLEFPINGSVVEREDVPAH-----LVKDIRNF 211
Db      352 QVYQQRRLRYVAFKQAPRCCK-----GHDDIPHEWGKALMEKWEAFL 393
QY      212 QVSPERF-SMLVAGK-----GNVKSMTSPKMSVVIYDILDSMQLRQEM 257
Db      394 SVKPOLFEKQIILMGODILKANYLKLLEGISVLAQAKSTMAKCTIKERIDISRYKKMDTV 453
QY      258 AIDQOSLGMRCQKSM 272
Db      454 QNFKSHGFCGDIIDM 468

```

```

Search completed: July 24, 2003, 12:58:01
Job time : 16.1534 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:27:29 ; Search time 41.1334 Seconds

(without alignments)
1234.824 Million cell updates/sec

Title: US-09-884-319A-4

Perfect score: 1645
Sequence: 1 KKGKTEQDGYKPTNKHFT.....DTLFEQKYVTLDSASFLLCSC 320

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_19Jun03:*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1645	100.0	320	18	AAW19990
2	1645	100.0	320	21	AAW37795
3	1343.5	81.7	950	20	AAV33298
4	1343.5	81.7	950	23	ABB81194
5	1343.5	81.7	950	24	ABR47911
6	1343.5	81.7	950	24	ABR00162
7	1057.5	64.3	316	21	AAW43502
8	921	56.0	260	23	ABJ05503
9	921	56.0	260	24	ABJ18424

10	871	52.9	278	17	AAW90544
11	869	52.8	783	20	AAW60344
12	747	45.4	188	22	AAW63253
13	513.5	31.2	189	22	ABW03204
14	513.5	31.2	189	24	ABU12498
15	159.5	9.7	509	22	AAW80430
16	159.5	9.7	509	22	AAW62173
17	159.5	9.7	509	22	AAW80382
18	159.5	9.7	509	23	ABW63574
19	159.5	9.7	509	23	ABW65272
20	159.5	9.7	509	24	ABW48117
21	159.5	9.7	509	24	ABW00285
22	159.5	9.7	554	22	AAW80407
23	159.5	9.7	554	23	ABW65271
24	159.5	9.7	554	24	ABW48118
25	159.5	9.7	554	24	ABW00286
26	159.5	9.7	594	22	AAW93619
27	147.5	9.0	431	21	AAW59009
28	147.5	9.0	465	21	AAW07747
29	117.5	7.1	198	24	ABW70859
30	105.5	6.4	1182	22	ABW48165
31	100	6.1	489	23	ABW54009
32	97	5.9	499	12	AAW11616
33	96	5.8	434	23	ABW55276
34	95.5	5.8	504	20	AAW93429
35	95.5	5.8	980	21	ABW18294
36	95.5	5.8	1818	23	ABW73779
37	95	5.8	480	7	ABW60263
38	95	5.8	499	20	AAW49569
39	95	5.8	510	21	AAW58148
40	94.5	5.7	1053	22	ABW11415
41	94	5.7	499	20	AAW49563
42	94	5.7	505	20	AAW32826
43	94	5.7	505	20	AAW32827
44	93.5	5.7	433	21	AAW30050
45	93.5	5.7	481	21	AAW30049

ALIGNMENTS

RESULT 1	AAW19990	standard; Protein; 320 AA.
XX	AAW19990;	
XX	27-AUG-1997 (first entry)	
XX	Type I, p80 IL-1-receptor intracellular domain ligand.	
AC	IL; interleukin; receptor; ligand; screening assay; inhibitor;	
XX	IL-1 mediated response; inflammation; inflammatory; antibody;	
KW	Intracellular domain; CAMP; calcium activated neutral protease.	
XX	Homo sapiens.	
OS	WO9640907-A1.	
XX	19-DEC-1996.	
XX	06-MAY-1996; 96WO-US06363.	
XX	07-JUN-1995; 95US-0487942.	
PR	(GENE) GENETICS INST INC.	
XX	Graham J, Lin L;	
XX	WPI: 1997-052315/05.	
DR	N-PSDB; AAT71217.	
XX	Interleukin-1 receptor intracellular ligand proteins and related DNA	

puG4-5-CDK-BP clon
Human normal blad
Human breast cance
Human musculokete
Novel human muscul
Gene #12 associate
Human gene 18-enco
Secreted protein e
Human albumin fusi
Human albumin fusi
Human secreted pro
Human gene 152 enc
Secreted protein e
Human albumin fusi
Human secreted pro
Human gene 152 enc
Human polypeptide,
Breast and ovarian
A human cancer-ass
Human adipocyte Se
Human testis-speci
Lactococcus lactis
Modified human hep
Lactococcus lactis
A. thaliana EL3 pr
Plasmodium falcipla
Candida albicans e
Sequence of new pr
Human heparin cofa
Lung cancer associ
Human phosphorilas
Human coagulation
Heparin cofactor I
Heparin cofactor I
Arabidopsis thalia
Arabidopsis thalia

PT - used to identify inhibitors of the proteins for treatment of
PT inflammation

PS Claim 3; Page 32-33; 54pp; English.

CC AAW1990 represents a protein that has interleukin-1 receptor (IL-1-R)
CC intracellular ligand activity. IL-1-R intracellular ligand proteins
CC are used to screen for agents (e.g. antibodies) that are capable of
CC inhibiting or blocking the binding of an IL-1-R intracellular ligand
CC to the intracellular domain of IL-1-R, i.e. inhibitors of IL-1
CC activity. Such agents can be used to treat inflammatory conditions.

XX Sequence 320 AA:

Query Match 100.0%; Score 1645; DB 18; Length 320;
Best Local Similarity 100.0%; Pred. No. 3e-164;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGKTEDDGYOKPNNKHTOSPKSVADLGSFEGKRRLLITAPKAKNNMYVOQRDEY 60
DB 1 KKGKTEDDGYOKPNNKHTOSPKSVADLGSFEGKRRLLITAPKAKNNMYVOQRDEY 60
QY 61 LESFCKMATRKISVITITGPNVNSMTMKIDHFDLNEKPMRVVDDDLVDQRLISELRKEX 120
DB 61 LESFCKMATRKISVITITGPNVNSMTMKIDHFDLNEKPMRVVDDDLVDQRLISELRKEX 120
QY 121 GMTYNDFFMVLTDVLDLRVQYVEVPIITMKSVPDLIDTQSRIDMEKOKKEGIVCKEEVG 180
DB 121 GMTYNDFFMVLTDVLDLRVQYVEVPIITMKSVPDLIDTQSRIDMEKOKKEGIVCKEEVG 180
QY 181 GYLELFPINGSSVVEREDVPAHLVDINRYFVSPEYSMLLVGDGNVKSYPSPMSM 240
DB 181 GYLELFPINGSSVVEREDVPAHLVDINRYFVSPEYSMLLVGDGNVKSYPSPMSM 240
QY 241 VIVYDLIDSMOLRQEMAIQOSLGMRCOKMSQMAVTTVTCTDMVTRMTTIVIRVITM 300
DB 241 VIVYDLIDSMOLRQEMAIQOSLGMRCOKMSQMAVTTVTCTDMVTRMTTIVIRVITM 300
QY 301 DTLTEQKYVTLDSASFCLSC 320
DB 301 DTLTEQKYVTLDSASFCLSC 320

RESULT 2
AAB37795
ID AAB37795 standard; Protein; 320 AA.

XX AAB37795;

DT 23-FEB-2001 (first entry)

XX Human interleukin-1 receptor intracellular ligand protein #2.

XX Human; interleukin-1; IL-1; IL-1alpha; IL-1beta; IL-1 receptor;
XX antiinflammatory; haemostatic; antibacterial; immunosuppressive;
XX immunomodulator; cardiac; cytostatic; neuroprotective; respiratory;
XX inflammation; infection; sepsis; cachexia; autoimmune disorder;
XX cardiovascular disorder; chronic myelogenous leukaemia;
XX multiple sclerosis; inflammatory bowel disease; Crohn's disease.

OS Homo sapiens.

XX WO200064479-A1.

PD 02-NOV-2000.

PF 26-APR-2000; 2000MO-US11700.

XX 27-APR-1999; 99US-0301274.

XX (ANTI-) ANTIBODY SYSTEMS INC.

XX Fredeking TM, Ignatyev GM;

XX WPI; 2000-679646/66.

XX Novel compositions comprising tetracycline or tetracycline-like
PT compounds for the treatment and/or prevention of acute inflammatory
PT responses and diseases, e.g. septic shock and immune complex-induced
PT colitis -

PS Disclosure; Page 156-157; 183pp; English.

CC The present sequence is given in a specification relating to novel
CC compositions and methods containing tetracycline or tetracycline-like
CC compounds for treating and/or preventing acute inflammatory responses and
CC diseases. Such diseases include acute inflammatory conditions associated
CC with viral haemorrhagic diseases (including diseases caused by
CC Bunyaviridae, Flaviviridae, Flaviviridae or Arenaviridae viruses),
CC parasitic diseases, bacterial infections, sepsis, cachexia, autoimmune
CC disorders, acute cardiovascular events, chronic myelogenous leukaemia and
CC transplanted bone marrow-induced graft-versus-host disease, septic shock,
CC immune complex-induced colitis, cerebrospinal fluid inflammation,
CC multiple sclerosis, inflammatory responses associated with trauma,
CC systemic inflammatory response syndrome (SIRS), adult respiratory
CC distress syndrome (ARDS), acute liver failure, inflammatory bowel disease
CC and Crohn's disease.

XX Sequence 320 AA:

Query Match 100.0%; Score 1645; DB 21; Length 320;
Best Local Similarity 100.0%; Pred. No. 3e-164;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKGKTEDDGYOKPNNKHTOSPKSVADLGSFEGKRRLLITAPKAKNNMYVOQRDEY 60
DB 1 KKGKTEDDGYOKPNNKHTOSPKSVADLGSFEGKRRLLITAPKAKNNMYVOQRDEY 60
QY 61 LESFCKMATRKISVITITGPNVNSMTMKIDHFDLNEKPMRVVDDDLVDQRLISELRKEX 120
DB 61 LESFCKMATRKISVITITGPNVNSMTMKIDHFDLNEKPMRVVDDDLVDQRLISELRKEX 120
QY 121 GMTYNDFFMVLTDVLDLRVQYVEVPIITMKSVPDLIDTQSRIDMEKOKKEGIVCKEEVG 180
DB 121 GMTYNDFFMVLTDVLDLRVQYVEVPIITMKSVPDLIDTQSRIDMEKOKKEGIVCKEEVG 180
QY 181 GYLELFPINGSSVVEREDVPAHLVDINRYFVSPEYSMLLVGDGNVKSYPSPMSM 240
DB 181 GYLELFPINGSSVVEREDVPAHLVDINRYFVSPEYSMLLVGDGNVKSYPSPMSM 240
QY 241 VIVYDLIDSMOLRQEMAIQOSLGMRCOKMSQMAVTTVTCTDMVTRMTTIVIRVITM 300
DB 241 VIVYDLIDSMOLRQEMAIQOSLGMRCOKMSQMAVTTVTCTDMVTRMTTIVIRVITM 300
QY 301 DTLTEQKYVTLDSASFCLSC 320
DB 301 DTLTEQKYVTLDSASFCLSC 320

RESULT 3
AAV33298
ID AAV33298 standard; Protein; 950 AA.

XX AAV33298;

DT 26-NOV-1999 (first entry)

XX Human membrane spanning protein MSP-5.

XX Membrane spanning protein; treatment; diagnosis; neoplastic disorder;
XX prevention; human; immunological disorder; reproductive disorder; MSP-5.

OS Homo sapiens.

XX WO9946380-A2.

PD 16-SEP-1999.
XX
PF 09-MAR-1999; 99MO-US05073.
XX
PR 13-MAR-1998; 98US-0039064.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Bandman O, Lal P, Hillman JL, Yue H, Corley NC;
PI Guegler KJ, Kaser MR, Baughn MR, Shah P;
XX
XX WPI; 1999-551409/46.
DR N-PSDB; AAZ09839, AAZ09840.
XX
PT New human membrane spanning proteins used to, e.g. prevent and treat
PT neoplastic disorders -
XX
PS Claim 1; Page 73-76; 81pp; English.
XX
CC This invention describes novel human membrane spanning proteins (MSPs),
CC and the polynucleotides encoding them. The products of the invention are
CC used to diagnose, prevent and treat neoplastic, immunological and
CC reproductive disorders. This sequence represents the human membrane
CC spanning protein MSP-5.
XX
SQ Sequence 950 AA;

Query Match 81.7%; Score 1343.5; DB 20; Length 950;
Best Local Similarity 81.4%; Pred. No. 8.4e-132;
Matches 267; Conservative 0; Mismatches 0; Indels 61; Gaps 1;

QY 1 KKGKTEODGYOKPTKHFQSPKSVADLLGSEFKRRLLITAPKAENMYVQORDEX 60
Db 586 KKGKTEODGYOKPTKHFQSPKSVADLLGSEFKRRLLITAPKAENMYVQORDEX 645
QY 61 LBSFCMAIRKISVITIFGPNVNSTKIDHFQDNEKPMRVNDEDLVDRLISELREK 120
Db 646 LBSFCMAIRKISVITIFGPNVNSTKIDHFQDNEKPMRVNDEDLVDRLISELREK 705
QY 121 GMTYNDFPFWLTVDLRVKQYEVPTITMKSVPFLIDTFOSRIKDMKKKEGIVCK 176
Db 706 GMTYNDFPFWLTVDLRVKQYEVPTITMKSVPFLIDTFOSRIKDMKKKEGIVCKEDKK 765
QY 177 -----EEV 179
Db 766 QSLNLFSPFRMRRLIVISAPNDEDMAYSQQLSALSGAQNGLRHITITLKLGVGEV 825
QY 180 GGVLELFPINGSSVVEREDVPAHLVKDIRNYFOVSPETYSMLLVGDKGNKSWYPSPPMS 239
Db 826 GGVLELFPINGSSVVEREDVPAHLVKDIRNYFOVSPETYSMLLVGDKGNKSWYPSPPMS 885
QY 240 MVTIVYDLIDSMOLRQEMAIQOSLGMRK 267
Db 886 MVTIVYDLIDSMOLRQEMAIQOSLGMRK 913

RESULT 4
ID ABB81194 standard; protein; 950 AA.
XX
XX ABB81194;
XX
XX 25-NOV-2002 (first entry)
XX
XX Human membrane spanning protein, MSP-5.
XX
XX Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;
XX membrane spanning protein; MSP-5; human.
OS
XX Homo sapiens.
XX
XX W0200262839-A2.
XX

PD 15-AUG-2002.
XX
XX 05-FEB-2002; 2002MO-EP01327.
PF
XX
XX 07-FEB-2001; 2001EP-0200439.
PR
XX
XX (UYMA-) UNIV MASTRICHT.
PA
XX
PI Daemen MAP, Cleutjens CBJM, Zaman GJR;
XX
XX WPI; 2002-643400/69.
DR N-PSDB; ABQ79519.
XX
PT Use of a polynucleotide differentially expressed in ruptured and stable
PT atherosclerotic plaques as a marker for atherosclerosis, useful in
PT treating, diagnosing or preventing atherosclerosis -
XX
XX Claim 1; Page 37-41; 44pp; English.
XX
PS The invention relates to the use of a polynucleotide differentially
XX expressed in ruptured and stable atherosclerotic plaques as a marker
XX for atherosclerosis, where the polynucleotides can be selected from the
XX sequences shown in ABQ79517-19. The polynucleotides are useful as a
XX marker of atherosclerosis, which may be used: (i) in the diagnosis,
XX prevention and treatment of atherosclerosis; (ii) as serum/plasma markers
XX to screen patients at risk for plaque instability to evaluate the effects
XX of other treatments; (iii) in the preparation of vector molecules for the
XX expression of the encoded protein in host cells; and (iv) in the
XX identification of functional targets or analogues of the gene. The
XX polynucleotides, the encoded proteins or antibodies against the proteins
XX may be used to target other therapeutics to an unstable plaque.
XX Modulation of the expression of the polynucleotide can increase plaque
XX stability and therefore inhibit the progression of atherosclerotic
XX cardiovascular disease. Modulators may be used to prepare pharmaceuticals
XX for atherosclerotic disorders. The present sequence represents a membrane
XX spanning protein, MSP-5.
XX
SQ Sequence 950 AA;

Query Match 81.7%; Score 1343.5; DB 23; Length 950;
Best Local Similarity 81.4%; Pred. No. 8.4e-132;
Matches 267; Conservative 0; Mismatches 0; Indels 61; Gaps 1;

QY 1 KKGKTEODGYOKPTKHFQSPKSVADLLGSEFKRRLLITAPKAENMYVQORDEX 60
Db 586 KKGKTEODGYOKPTKHFQSPKSVADLLGSEFKRRLLITAPKAENMYVQORDEX 645
QY 61 LBSFCMAIRKISVITIFGPNVNSTKIDHFQDNEKPMRVNDEDLVDRLISELREK 120
Db 646 LBSFCMAIRKISVITIFGPNVNSTKIDHFQDNEKPMRVNDEDLVDRLISELREK 705
QY 121 GMTYNDFPFWLTVDLRVKQYEVPTITMKSVPFLIDTFOSRIKDMKKKEGIVCK 176
Db 706 GMTYNDFPFWLTVDLRVKQYEVPTITMKSVPFLIDTFOSRIKDMKKKEGIVCKEDKK 765
QY 177 -----EEV 179
Db 766 QSLNLFSPFRMRRLIVISAPNDEDMAYSQQLSALSGAQNGLRHITITLKLGVGEV 825
QY 180 GGVLELFPINGSSVVEREDVPAHLVKDIRNYFOVSPETYSMLLVGDKGNKSWYPSPPMS 239
Db 826 GGVLELFPINGSSVVEREDVPAHLVKDIRNYFOVSPETYSMLLVGDKGNKSWYPSPPMS 885
QY 240 MVTIVYDLIDSMOLRQEMAIQOSLGMRK 267
Db 886 MVTIVYDLIDSMOLRQEMAIQOSLGMRK 913

RESULT 5
ID ABR47911 standard; protein; 950 AA.
XX
XX ABR47911;
AC

Query Match	Best Local Similarity	81.7%	Score 1343.5	DB 24	Length 950
Matches 267	Conservative	0	Mismatches 132	Indels 61	Gaps 1
1	KKGKTEDDGQYKPNKHFHTSPKRSVADLLSFEKRRLLITAPKAENMNVVOORDEY	60			
586	KKGKTEDDGQYKPNKHFHTSPKRSVADLLSFEKRRLLITAPKAENMNVVOORDEY	645			
61	LESECKMARRISVITIEGPVNNSTPMKIDHFOIDNEKPMRVVDEDDLVORLISELRKEY	120			
646	LESECKMARRISVITIEGPVNNSTPMKIDHFOIDNEKPMRVVDEDDLVORLISELRKEY	705			
121	GMTNDEFMVLTVDLRAVKQYIEVPIITMKSVDLIDTFOSRIKDEKQKKEGIVCK----	176			

Db	706	GMRYNDFPVLTDVLRVQYYEVEITTKSVFDDLTDFQSIKMEKQKKEGIVCKEDDK	765
Oy	177	-----EYV 179	111
Db	766	QSLNFELSRFRMRRLIVISAPNDEDMAYSQQLSALSGQACNFGIRHITILKILGVGEV	825
Oy	180	GGYLELEFPINGSVVEREDVPALHVKQDIRNFQVSPFEFSMLVGRKGNKSWPSPMWS	239
Db	826	GGYLELEFPINGSVVEREDVPALHVKDIRNFQVSPFEFSMLVGRKGNKSWPSPMWS	885
Oy	240	MVIVYDIDSMOLRQEMAIQOOSLMRC 267	111
Db	886	MVIVYDIDSMOLRQEMAIQOOSLMRC 913	111
RESULT 6			
ABR00162			
ID	ABR00162	standard; Protein; 950 AA.	
XX	ABR00162;		
AC			
XX	03-APR-2003	(first entry)	
DT			
XX			
DE	Human gene 152	encoded secreted protein HSKDA27, SEQ ID NO:451.	
XX			
KW	Human; secreted protein; digestive disorder; gastrointestinal disorder;		
KW	mouth; oesophagus; stomach; small intestine; large intestine; liver;		
KW	biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;		
KW	immune disorder; inflammation; infection; wound healing; drug screening;		
KW	chromosome identification; chromosome mapping; cytostatic;		
KW	antiinflammatory; immunosuppressive; vulnerary; gene therapy.		
XX			
OS	Homo sapiens.		
PN	WO200276488-A1.		
XX			
PD	03-OCT-2002.		
XX			
PE	19-MAR-2002; 2002WO-US08276.		
XX			
PR	21-MAR-2001; 2001US-277340P.		
PR	19-JUL-2001; 2001US-306171P.		
PR	13-NOV-2001; 2001US-331287P.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Rosen CA, Ruben SM;		
XX			
DR	WPI: 2003-029900/02.		
DR	N-PSDB: ABZ71341.		
XX			
PT	New human secreted proteins and nucleic acids, useful for detecting,		
PT	preventing, diagnosing, prognosticating, treating and/or ameliorating		
PT	e.g. gastrointestinal diseases and disorders, or cancers -		
XX			
XX	Claim 13; Page 1038-1041; 1216pp; English.		
CC	ABZ71190-ABZ71478	represent cDNAs corresponding to 178 human secreted	
CC	protein genes, and ABR000011-ABR000299	represent the proteins they encode.	
CC	ABZ71479-ABZ71540	represent human secreted protein genomic fragments. The	
CC	invention also encompasses antibodies specific for the secreted proteins,		
CC	the use of the secreted proteins in drug screening, and recombinant		
CC	vectors and host cells comprising a nucleic acid of the invention. The		
CC	secreted proteins, nucleic acids encoding them, antibodies or antibody		
CC	fragments specific for the secreted proteins, and modulators of protein		
CC	activity are useful for diagnosing, treating, ameliorating or preventing		
CC	digestive disorders. Such conditions include disorders of the mouth,		
CC	oesophagus, stomach, small intestine, large intestine, liver, biliary		
CC	tract and pancreas, and include cancers of these organs and tissues. The		
CC	secreted proteins and their nucleic acids may also be used in the		
CC	treatment of immune disorders, inflammation, infection,		
CC	hyperproliferative disorders, and to promote wound healing. Nucleic acids		

CC of the invention may be used for chromosome identification, chromosome
CC mapping, in gene therapy, for identifying individuals from minute
CC biological samples, as hybridisation probes, and as molecular weight
CC markers. The present sequence represents a human secreted protein of the
CC invention.

XX Sequence 950 AA:

Query Match 81.7%; Score 1343.5; DB 24; Length 950;
Best Local Similarity 81.4%; Pred. No. 8.4e-132;
Matches 267; Conservative 0; Mismatches 0; Indels 61; Gaps 1;

QY 1 KKGGTEODGYQKPTKKHFTQSPKSVADLLGSFEKRRLLITTAERNNMYVQGRDEX 60
DB 586 KKGGTEODGYQKPTKKHFTQSPKSVADLLGSFEKRRLLITTAERNNMYVQGRDEX 645
QY 61 LESFCMAATRKISVITTFGPVNSTKIDHFOLDNKPMPVYDDELDVQRLISELRKEY 120
DB 646 LESFCMAATRKISVITTFGPVNSTKIDHFOLDNKPMPVYDDELDVQRLISELRKEY 705
QY 121 GMTYNDFFVYLDVLRVQYEVPTTKMSVFDLIDTFQSRIDKMEKQKKEGIVCK--- 176
DB 706 GMTYNDFFVYLDVLRVQYEVPTTKMSVFDLIDTFQSRIDKMEKQKKEGIVCKEDDK 765
QY 177 -----EYV 179
DB 766 QSLLENFLSRFRMRRLVTSAPNDEDMAYSQLSQAQNGFLRHITILKLLGVGEY 825
QY 180 GGVLELFPINGSSVVEREDVPAHLVKDIRNYFQVSPFYFSLVGGKGNVKSYPSPMWS 239
DB 826 GGVLELFPINGSSVVEREDVPAHLVKDIRNYFQVSPFYFSLVGGKGNVKSYPSPMWS 885
QY 240 MVIYVDLIDSMQLRQEMAIQOOSLGMRG 267
DB 886 MVIYVDLIDSMQLRQEMAIQOOSLGMRG 913

RESULT 7
AAB43502
ID AAB43502 standard; Protein; 316 AA.

AC AAB43502;

DT 08-FEB-2001 (first entry)

XX Human cancer associated protein sequence SEQ ID NO:947.

XX Human: cancer associated gene; cancer antigen; detection; cancer;
KM diagnosis; cytostatic; proliferative; vulnerrary; immunomodulator;
KM antidiabetic; antitumoric; antirheumatic; antiarthritic; antiviral;
KM antiinflammatory; antihypertoid; antiallergic; antibacterial; coagulant;
KM dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KM vasotropic; antiproliferic; antiangiogenic; gene therapy; inflammation;
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM allergic reaction; graft versus host disease; organ rejection;
KM haemostatic; thrombolytic; cardiovascular disorder; infection;
KM neurological disease; drug screening.

OS Homo sapiens.

XX WO200053350-A1.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05882.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

XX WPI; 2000-587533/55.

DR N-PSDB; AAC77711.

XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -

PS Claim 11; Page 1512-1514; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerrary; immunomodulator;
CC antidiabetic; antitumoric; antirheumatic; antiarthritic;
CC antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; coagulant; thrombolytic; coagulant;
CC nootropic; vasotropic; antiproliferic; antiangiogenic; gene therapy;
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC79457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.

XX Sequence 316 AA:

Query Match 64.3%; Score 1057.5; DB 21; Length 316;
Best Local Similarity 77.7%; Pred. No. 2e-102;
Matches 212; Conservative 0; Mismatches 0; Indels 61; Gaps 1;

QY 56 QREYLESFCMAATRKISVITTFGPVNSTKIDHFOLDNKPMPVYDDELDVQRLISE 115
DB 7 QREYLESFCMAATRKISVITTFGPVNSTKIDHFOLDNKPMPVYDDELDVQRLISE 66
QY 116 LRKEYGMTYNDFFVYLDVLRVQYEVPTTKMSVFDLIDTFQSRIDKMEKQKKEGIVC 175
DB 67 LRKEYGMTYNDFFVYLDVLRVQYEVPTTKMSVFDLIDTFQSRIDKMEKQKKEGIVC 126
QY 176 K----- 176
DB 127 KEDKQSLLENFLSRFRMRRLVTSAPNDEDMAYSQLSQAQNGFLRHITILKLLG 186
QY 177 --EEVGVLLELFPINGSSVVEREDVPAHLVKDIRNYFQVSPFYFSLVGGKGNVKSYP 234
DB 187 VGEVEGVLELFPINGSSVVEREDVPAHLVKDIRNYFQVSPFYFSLVGGKGNVKSYP 246
QY 235 SPWMSVVIYVDLIDSMQLRQEMAIQOOSLGMRG 267
DB 247 SPWMSVVIYVDLIDSMQLRQEMAIQOOSLGMRG 279

RESULT 8
ABJ05503

ID ABJ05503 standard; Protein; 260 AA.

XX ABJ05503;

XX 14-NOV-2002 (first entry)

DE Human breast cancer associated polypeptide SEQ ID NO: 263.

XX Human: breast specific gene; breast specific protein; breast cancer;
KM gene therapy; cytostatic.

XX Homo sapiens.

XX WO200264611-A1.

PD 22-AUG-2002.
XX
XX 12-FEB-2002: 2002MO-US04197.
XX
XX 13-FEB-2001: 2001US-268292P.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Salceda S, Macina RA, Hu P, Recipon H, Karra K, Caferkey R;
PI Sun Y, Liu C;
XX
XX WPI: 2002-657582/70.
XX
XX New breast specific nucleic acids and proteins, useful for identifying,
PT diagnosing, monitoring, staging, imaging, and treating breast cancer
PT and non-cancerous disease states in breast tissue, and in gene therapy
PT
XX
XX Claim 11; Page 340-341; 367pp; English.
XX
XX The present invention provides human breast specific coding sequences and
CC proteins. These can be used in the diagnosis and treatment of breast
CC cancer and non-cancerous diseases of the breast. The present sequence is
CC a polypeptide of the invention.
XX
XX Sequence 260 AA:
SQ
Query Match 56.0%; Score 921; DB 23; Length 260;
Best Local Similarity 96.8%; Pred. No. 3.4e-88;
Matches 180; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 KKGKTEODGYOKPTNKHFTQSPKRSVADLLGSFEKRRLLITAPKAENMYVOQRDEY 60
DB |||||
DB 38 KKGKTEODGYOKPTNKHFTQSPKRSVADLLGSFEKRRLLITAPKAENMYVOQRDEY 97
QY 61 LESFCMATRKISVITIFGPVNNSTMKIDHFDLNEKPKRVVDEDLVDQRLISELRKEY 120
DB |||||
DB 98 LESFCMATRKISVITIFGPVNNSTMKIDHFDLNEKPKRVVDEDLVDQRLISELRKEY 157
QY 121 GMTYNDFEVLVDVLRVQYEVPTMKSVFDLIDTFSRIKDMKOKKEGIVCKEY 180
DB |||||
DB 158 GMTYNDFEVLVDVLRVQYEVPTMKSVFDLIDTFSRIKDMKOKKEGIVCKEDK 217
QY 181 GVLELF 186
DB |||
DB 218 OSLENF 223
RESULT 9
ID AAR90544
ID AAR90544 standard; protein; 260 AA.
XX
XX AAR90544;
AC
XX
XX 30-JUN-2003 (first entry)
DT
XX
XX Breast specific related amino acid sequence SEQ ID No 233.
DE
XX
XX Cytostatic; BSP-agonist; BSP-antagonist; vaccine; gene therapy; cancer;
KW metastatic; breast cancer; breast specific; human.
XX
XX Homo sapiens.
OS
XX
XX WO20027232-A2.
PN
XX
XX 03-OCT-2002.
PD
XX
XX 21-NOV-2001; 2001MO-US43815.
PF
XX
XX 22-NOV-2000; 2000US-252509P.
PR
XX
XX (DIAD-) DIADEXUS INC.
PA
XX

PI Salceda S, Macina RA, Recipon H, Pluta J, Sun Y, Liu C;
XX
XX WPI: 2003-018927/01.
XX
XX New isolated nucleic acid molecule, useful for treating breast cancer,
PT and diagnosing or monitoring the presence of metastases of breast
PT cancer in a patient
XX
XX
XX Claim 11; Page 352-353; 377pp; English.
XX
XX The invention relates to a novel isolated nucleic acid molecule
CC comprising a sequence encoding a sequence comprising 11-1518 amino
CC acids; a sequence comprising 190-8144 bp; or a sequence that selectively
CC hybridises to, or having at least 60% identity with the 11-1518 amino
CC acid or 190-8144 nt sequence. The polypeptide and the nucleic acid are
CC useful for treating breast cancer, and diagnosing or monitoring the
CC presence of metastases of breast cancer in a patient. The polynucleotides
CC of the invention can be used to treat disorders by gene therapy. This
CC invention.
XX
XX Sequence 260 AA:
SQ
Query Match 56.0%; Score 921; DB 24; Length 260;
Best Local Similarity 96.8%; Pred. No. 3.4e-88;
Matches 180; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 KKGKTEODGYOKPTNKHFTQSPKRSVADLLGSFEKRRLLITAPKAENMYVOQRDEY 60
DB |||||
DB 38 KKGKTEODGYOKPTNKHFTQSPKRSVADLLGSFEKRRLLITAPKAENMYVOQRDEY 97
QY 61 LESFCMATRKISVITIFGPVNNSTMKIDHFDLNEKPKRVVDEDLVDQRLISELRKEY 120
DB |||||
DB 98 LESFCMATRKISVITIFGPVNNSTMKIDHFDLNEKPKRVVDEDLVDQRLISELRKEY 157
QY 121 GMTYNDFEVLVDVLRVQYEVPTMKSVFDLIDTFSRIKDMKOKKEGIVCKEY 180
DB |||||
DB 158 GMTYNDFEVLVDVLRVQYEVPTMKSVFDLIDTFSRIKDMKOKKEGIVCKEDK 217
QY 181 GVLELF 186
DB |||
DB 218 OSLENF 223
RESULT 10
ID AAR90544
ID AAR90544 standard; protein; 278 AA.
XX
XX AAR90544;
AC
XX
XX 08-AUG-1996 (first entry)
DT
XX
XX pJG4-5-CDK-BP clone #125 derived CDK4 binding protein.
DE
XX
XX Cell cycle; CDK4; regulation; G1 phase; proliferation; tumourigenesis;
KW cyclin dependent kinase; differentiation; CDK4 inhibitor; agonist;
XX
XX antagonist.
XX
XX Synthetic.
OS
XX
XX WO9533819-A2.
PN
XX
XX 14-DEC-1995.
PD
XX
XX 02-JUN-1995; 95MO-US07113.
PF
XX
XX 02-JUN-1994; 94US-0253155.
PR
XX
XX (MITO-) MITOFIX INC.
PA
XX
XX Draetta G, Gyuris J;
PI
XX
XX WPI: 1996-040227/04.
DR

DR N-PSDB; AAT12173.
 XX Cyclicin-dependent kinase-4 binding protein - used in the isolation of
 PT (ant)agonists of cell cycle regulation.
 XX
 PS Claim 1; Page 81-82; 115pp; English.
 CC AAR0533-R90556 are cyclin dependent kinase 4 (CDK4) binding proteins.
 CC encoded by clones of the plasmid pUG4-5-CDKBP. CDK4 binding proteins
 CC (CDK4-BP) may be used in an assay for screening test compounds as
 CC inhibitors of CDK/CDK4-BP interaction. The complexes formed by CDK4
 CC and D-type cyclins are strongly implicated in the control of the early
 CC G1 phase of the cell cycle and are strong candidates for controlling
 CC and/or preventing tumourigenesis and the onset of cancer. Nucleic acids
 CC encoding CDK4-BP or fragments of these may be used as probes/primers
 CC to diagnose the presence or absence of genetic lesions in a gene
 CC encoding 1 of the 24 CDK4-BP, and hence to diagnose the risk for a
 CC subject of developing a cell-proliferation associated disorder (e.g.
 CC cancer).
 XX
 SQ Sequence 278 AA:
 Query Match 52.9%; Score 871; DB 17; Length 278;
 Best Local Similarity 69.1%; Pred. No. 6.8e-83;
 Matches 183; Conservative 3; Mismatches 13; Indels 66; Gaps 3;
 QY 1 KKGKTEODGYOKPNNKHFTOSPKKSVADLLG-SFEGRRLLLITAPKAENMYVOORDE 59
 DB 13 KKGKTEODGYOKPNNKHFTOSPKKSVADLLG-SFEGRRLLLITAPKAENMYVOORDE 68
 QY 60 YLESCKKATRKISVITTFGPNVNSTMKIDHFOLDNEKPMRVVDEDLVDQRLISELRKE 119
 DB 69 YLESCKKATRKISVITTFGPNVNSTMKIDHFOLDNEKPMRVVDEDLVDQRLISELRKE 128
 QY 120 GMTYNDFEWMVLTVDLRAVKQYEVPIITMKSVFIDLIDFOSRIKMEKOKKEGIYCK-- 176
 DB 129 YGMTYNDFEWMVLTVDLRAVKQYEVPIITMKSVFIDLIDFOSRIKMEKOKKEGIYCKEDK 188
 QY 177 -----EE 178
 DB 189 KQSLNPLSRFRMRRLLVISAPNDEDMAYSOQLSALSGQACNFGRLRITILKLVGGE 248
 QY 179 VGVLELFPINGSSVVEREDVPAHL 203
 DB 249 VGVLELFPINGSSVVEREDVPAHL 273
 RESULT 11
 AAY60344
 ID AAY60344 standard; Protein; 783 AA.
 XX
 AC AAY60344;
 XX
 DT 31-JAN-2000 (first entry)
 XX
 DE Human normal bladder tissue EST encoded protein 16.
 XX
 KW Human; bladder; treatment; EST; expressed sequence tag; cytostatic;
 KW cancer; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN DE19818620-A1.
 PD 28-OCT-1999.
 XX
 PF 21-APR-1998; 98DE-1018620.
 XX
 PR 21-APR-1998; 98DE-1018620.
 XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX
 DR WPI; 1999-602416/52.
 DR N-PSDB; AA242150.
 XX
 PT New polypeptides and their nucleic acids, useful for treatment of
 PT bladder tumour and identification of therapeutic agents -
 XX
 PS Claim 23; Page 254; 366pp; German.
 CC This invention describes novel polypeptide fragment sequences (I) and
 CC their encoding nucleic acids (II) which are highly expressed in normal
 CC bladder tissue and have cytostatic activity. (II) are used for
 CC recombinant expression of (I) and to isolate complete genes. (I) are
 CC used to identify agents suitable for the treatment of bladder tumours,
 CC to directly treat this form of cancer (including expression from gene
 CC therapy vectors), or are used in a preparation for cancer treatment. (I)
 CC is also used for the generation of specific antibodies. (II) are
 CC identified by assembling ESTs (expressed sequence tags) from a
 CC particular tissue type before comparison of expression patterns. This
 CC allows a significantly longer fragment of the gene to be revealed, and
 CC therefore reduces the number of failures because of ESTs from different
 CC libraries representing different parts of the same unknown gene
 CC distorting the estimated frequency of occurrence in a particular tissue.
 CC AAY60329-Y60591 represent protein fragments encoded by the human normal
 CC bladder tissue cDNA library derived EST fragments represented in
 CC AA242122-242248.
 XX
 SQ Sequence 783 AA:
 Query Match 52.8%; Score 869; DB 20; Length 783;
 Best Local Similarity 98.8%; Pred. No. 5.4e-82;
 Matches 168; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKGKTEODGYOKPNNKHFTOSPKKSVADLLGSFEGKRLLLITAPKAENMYVOORDEY 60
 DB 604 KKGKTEODGYOKPNNKHFTOSPKKSVADLLGSFEGKRLLLITAPKAENMYVOORDEY 663
 QY 61 LESFCMATRKISVITTFGPNVNSTMKIDHFOLDNEKPMRVVDEDLVDQRLISELRKEY 120
 DB 664 LESFCMATRKISVITTFGPNVNSTMKIDHFOLDNEKPMRVVDEDLVDQRLISELRKEY 723
 QY 121 GMTYNDFEWMVLTVDLRAVKQYEVPIITMKSVFIDLIDFOSRIKMEKOKK 170
 DB 724 GMTYNDFEWMVLTVDLRAVKQYEVPIITMKSVFIDLIDFOSRIKMEKOKR 773
 RESULT 12
 AAB63253
 ID AAB63253 standard; Protein; 188 AA.
 XX
 AC AAB63253;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE Human breast cancer associated antigen protein sequence SEQ ID NO:615.
 XX
 KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
 KW cancer associated antigen; cytostatic; cancer vaccine.
 XX
 OS Homo sapiens.
 XX
 PN WO200073801-A2.
 PD 07-DEC-2000.
 XX
 PF 26-MAY-2000; 2000MO-US14749.
 XX
 PR 28-MAY-1999; 99US-0136526.
 XX
 PR 10-SEP-1999; 99US-0153454.
 XX
 PA (LUDWIG-) LUDWIG INST CANCER RES.
 XX
 PI Obata Y;

XX WPI: 2001-025274/03.
DR Nucleic acids encoding breast, gastric and prostate cancer associated
XX antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
PS Example 1: Page 483-484; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterized by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
SQ Sequence 188 AA:

Query Match 45.4%; Score 747; DB 22; Length 188;
Best Local Similarity 100.0%; Freq. No. 4.2e-70;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKGKTEODGYQKPTNKHFTQSPKKSVDLGSFEGKRLLITAPKAENMYVOORDEY 60
DB 44 KKGKTEODGYQKPTNKHFTQSPKKSVDLGSFEGKRLLITAPKAENMYVOORDEY 103
OY 61 LESFCKMATRKISVITTFEGPVNNSTWKIDHFDLDNEKPMRVVDEDLVDQRLISELRKEY 120
DB 104 LESFCKMATRKISVITTFEGPVNNSTWKIDHFDLDNEKPMRVVDEDLVDQRLISELRKEY 163
OY 121 GMTYDFEYVLTVDLRYKQYEV 144
DB 164 GMTYDFEYVLTVDLRYKQYEV 187

RESULT 13
AAB03204
ID ABB03204 standard; Protein; 189 AA.
XX
XX ABB03204;
AC
XX
XX
DT 08-JAN-2002 (first entry)
XX
XX Human musculoskeletal system related polypeptide SEQ ID NO 1151.
DE
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
KW vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system.
XX
XX Homo sapiens.
OS
XX
XX WO200155367-A1.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01338.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217996.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226861.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227109.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.

[illegible]

CC	Query Match	31.2%	Score 513.5	DB 22	Length 189
CC	Best Local Similarity	42.6%	Pred. No. 1.6e-45		
CC	Matches 118	Conservative 12	Mismatches 30	Indels 117	Gaps 2
QY	21	OSPCKSVADLISFEBCRKRLITPAKKNENNYGQREVLSESCMKATRKISVITFGP	80		
DB	23	EDRKQSDNLFSEFRMRRLVISAANDDMAYSQQLSAGACNGLRHTITLKLIG-	81		
QY	81	VNNSTKIDHFQLDNEKPRKVVDEDLVDQRLISELRKREYGMVFYVLTVDLRYKQ	140		
DB	82	-----	81		
QY	141	YVEVPTKMSVFDLIDTFQSRIKDMEKOKKESTVCKEEVGVLELFPINGSSVEREDVP	2000		
DB	82	-----VGEVGVLELFPINGSSVEREDVP	1078		
QY	201	AHLVKIDIRNFQVSPFESMLLVGDKGNKSNYPSPMSMYIVVDLDSMQLRQEMAIQ	2666		
DB	108	AHLVKIDIRNFQVSPFESMLLVGDKGNKSNYPSPMSMYIVVDL-----	1555		
QY	261	QSLGMRCKMSMQAMVTITVTFTDTRVMTVIMRV	297		
DB	156	-----SMQAMVTITVTFTDTRVMTVIMRV	182		
RESULT 14					
ABU12498					
ID	ABU12498	standard; Protein; 189 AA.			
XX	ABU12498;				
XX	26-FEB-2003	(first entry)			
DE	Novel human musculoskeletal system antigen #118.				
XX	Musculoskeletal system antigen; cancer; metastasis;				
KM	re-vascularisation; thrombosis; arteriosclerosis; mineral content;				
KM	cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;				
KM	post-operative tissue repair; limb regeneration; neuronal growth;				
KM	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;				
KM	AIDS-related complex; chondrocyte growth; bone regeneration;				
KM	periodontal regeneration; tissue transport; bone graft; skin aging;				
KM	keratinocyte growth; hair loss; melanocyte growth; cell proliferation;				
KM	cell growth; organ transplant; cell differentiation; body height;				
KM	weight; hair colour; eye colour; skin; percentage of adipose tissue;				
KM	pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;				
KM	depression; tendency for violence; pain; reproductive capability;				
KM	hormone level; endocrine level; appetite; libido; memory; stress;				
KM	storage capability; fat content; lipid content; protein content;				
KM	carbohydrate content; vitamin content; cofactor content;				
XX	nutritional component.				
OS	Homo sapiens.				
XX					

US2002147140-A1.

10-OCT-2002.

17-JAN-2001; 2001US-0764877.

31-JAN-2000; 2000US-179065P.

04-FEB-2000; 2000US-180628P.

28-JUN-2000; 2000US-214886P.

07-JUL-2000; 2000US-216647P.

07-JUL-2000; 2000US-216880P.

11-JUL-2000; 2000US-217487P.

11-JUL-2000; 2000US-217496P.

14-JUL-2000; 2000US-218290P.

26-JUL-2000; 2000US-220963P.

26-JUL-2000; 2000US-220964P.

14-AUG-2000; 2000US-224518P.

14-AUG-2000; 2000US-224519P.

14-AUG-2000; 2000US-225267P.

14-AUG-2000; 2000US-225268P.

14-AUG-2000; 2000US-225270P.

14-AUG-2000; 2000US-225447P.

14-AUG-2000; 2000US-225757P.

14-AUG-2000; 2000US-225758P.

22-AUG-2000; 2000US-226868P.

30-AUG-2000; 2000US-228924P.

01-SEP-2000; 2000US-229287P.

01-SEP-2000; 2000US-229343P.

01-SEP-2000; 2000US-229344P.

01-SEP-2000; 2000US-229345P.

05-SEP-2000; 2000US-229509P.

05-SEP-2000; 2000US-229513P.

08-SEP-2000; 2000US-231413P.

21-SEP-2000; 2000US-234223P.

21-SEP-2000; 2000US-234274P.

25-SEP-2000; 2000US-234977P.

27-SEP-2000; 2000US-235834P.

29-SEP-2000; 2000US-236327P.

29-SEP-2000; 2000US-236367P.

29-SEP-2000; 2000US-236368P.

29-SEP-2000; 2000US-236369P.

29-SEP-2000; 2000US-236370P.

02-OCT-2000; 2000US-236802P.

02-OCT-2000; 2000US-237037P.

02-OCT-2000; 2000US-237038P.

02-OCT-2000; 2000US-237039P.

13-OCT-2000; 2000US-237040P.

13-OCT-2000; 2000US-239935P.

20-OCT-2000; 2000US-240960P.

20-OCT-2000; 2000US-241785P.

20-OCT-2000; 2000US-241809P.

01-NOV-2000; 2000US-244617P.

17-NOV-2000; 2000US-249299P.

08-DEC-2000; 2000US-251856P.

08-DEC-2000; 2000US-251868P.

08-DEC-2000; 2000US-251869P.

(ROSE/) ROSEN C A.

(RUBE/) RUBEN S M.

(BARA/) BARASH S C.

Rosen CA, Ruben SM, Barash SC;

WPI: 2003-128199/12.

N-PSDB: ABX57774.

Isolated nucleic acid molecules encoding musculoskeletal system associated polypeptides, useful for detecting disorders, e.g. cancer

Claim 11: SEQ ID NO 1151; 321pp; English.

The invention describes an isolated nucleic acid molecule comprising a sequence encoding musculoskeletal system associated polypeptides useful

for detecting disorders, e.g., cancer or cancer metastases, in animals or humans. The nucleic acid: stimulates re-vascularisation of ischemic tissues associated with conditions such as thrombosis, arteriosclerosis, and other cardiovascular conditions; treats wounds due to injuries, burns, post-operative tissue repair, and ulcers; stimulates angiogenesis and limb regeneration; stimulates neuronal growth; can treat and prevent neuronal damage occurring in certain disorders or neurodegenerative conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-related complex; stimulates chondrocyte growth, thus they can be used to enhance bone and periodontal regeneration and aid in tissue transports or bone grafts; prevents skin aging due to sunburn by stimulating keratinocyte growth; prevents hair loss, since FGF family members activate hair-forming cells and promotes melanocyte growth; stimulates growth and differentiation of hematopoietic cells and bone marrow cells when used in combination with other cytokines; maintains organs before transplantation or for supporting cell culture of primary tissues; induces tissue of mesodermal origin to differentiate in early CC embryos; increases or decreases the differentiation or proliferation of CC embryonic stem cells, besides, haematopoietic lineage; modulates CC mammalian characteristics, such as, body height, weight, hair colour, eye colour, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery); modulates mammalian metabolism; changes CC mammal's metal state or physical state by influencing biorhythms, CC circadian rhythms, depression, tendency for violence, tolerance for pain, CC reproductive capabilities, hormonal or endocrine levels, appetite, CC libido, memory, or stress; increases or decreases storage capabilities, CC fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components. This is the amino acid sequence of a CC novel human musculoskeletal system antigen.

Note: The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly CC from the US patent Office at

ftp.segdata.uspto.gov/sequence.html?docID=20020147140.

Sequence 189 AA:

Query Match 31.2%; Score 513.5; DB 24; Length 189;

Best Local Similarity 42.6%; Pred. No. 1.6e-45;

Matches 118; Conservative 12; Mismatches 30; Indels 117; Gaps 2;

21 OSPKSVADLGSFEGKRRLLITPKAENNNYQGRREYLESFCKMTRKISVTTCIP 80

23 EDKROSLNLFSEFRWRRLVTSAPNDEDMWYSQSLSGQACNFCRLHTTIKLG- 81

81 VNNSTKIDHFOLDNEKPMRVVDEDLVDQRLISELRREYGMTYNDFFMVLVDVLRVQ 140

82 ----- 81

141 YIEVPTMKSVFDLIDTQSRIMKEKOKKESIGCKEYGVLELFPINGSSVVEREDVP 200

82 -----VGEVGVLELFPINGSSVVEREDVP 107

201 AHLVKDIRNPVSPEYFMSMLYKGDGVKWSYPSPMASMYIVYDLDLSMLRQEMAIQ 260

108 AHLVKDIRNPVSPEYFMSMLYKGDGVKWSYPSPMASMYIVYDLD----- 155

261 QSLGRCOKMSQWAVTIVTTKDTFRMYTRMTYIVIKRV 297

156 -----SMQAMVITVTRKDFRMYTRMTYIVIKRV 182

RESULT 15

AAB80430

ID AAB80430 strand: peptide; 505 AA.

XX

AC AAB80430;

XX

DT 24-APR-2001 (first entry)

XX

DE Gene #12 associated peptide #1.

XX

KW Secreted protein; human; autoimmune; hyperproliferation;

KM cardiovascular; cerebrovascular; infection; food.

Query Match	100.0%;	Score 1294;	DB 6;	Length 700;
Best Local Similarity	100.0%;	Pred. No. 8.4e-97;		
Matches 251; Conservative	0;	Mismatches	0;	Indels 0;
				Gaps 0;

```

QY 1 KNEFLTNRARSRDPTFNLREVNLNFKLPPEGYILVPSTFEPNKGDCIRVFSEKKADY 60
DB 450 KNEFLTNRARSRDPTFNLREVNLNFKLPPEGYILVPSTFEPNKGDCIRVFSEKKADY 509
QY 61 QAVDEIANTANLEEDPISDDIDGFRRLFAQLAGDAEISAFELQTLIRRYLAKRODIX 120
DB 510 QAVDEIANTANLEEDPISDDIDGFRRLFAQLAGDAEISAFELQTLIRRYLAKRODIX 569
QY 121 DGFSEITCKIWMVMDSDSGSKLGKLEFYILMTKIOKTYREIDVDPSGTMSYEMRK 180
DB 570 DGFSEITCKIWMVMDSDSGSKLGKLEFYILMTKIOKTYREIDVDPSGTMSYEMRK 629
QY 181 ALEAGFKMPCOLHOVIYARFADDLIDFNFVRCVLRLETFKIFKOLDPENTGTIEL 240
DB 630 ALEAGFKMPCOLHOVIYARFADDLIDFNFVRCVLRLETFKIFKOLDPENTGTIEL 689
QY 241 DLISWLCFSVL 251
DB 690 DLISWLCFSVL 700

```

RESULT 2

```

ID 09YIC1 PRELIMINARY; PRT; 705 AA.
AC 09YIC1;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Quail calpain.
OS Coturnix coturnix (Common quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
NCBI_TaxID=9091;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Maeda Y.;
RT "cDNA sequence of quail skeletal muscle calpain.";
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB011080; BAA74564.1; -.
DR HSSP: P04574; IALV.
DR MEROPS: C02.003; -.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Protease_C2.
DR InterPro: IPR000169; SHPTC_acsite.
DR Pfam: PF01067; Calpain_III; 1.
DR Pfam: PF00036; efhand; 2.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00720; calpain_III; 1.
DR SMART: SM00230; CysPc; 1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00139; THIOI_PROTEASE_CYS; 1.
SQ SEQUENCE 705 AA; 80188 MW; 5C70634EB3AA93CD CRC64;

```

Query Match 61.7%; Score 798.5; DB 13; Length 705;
 Best Local Similarity 57.4%; Pred. No. 1.6e-56;
 Matches 144; Conservative 61; Mismatches 45; Indels 1; Gaps 1;

```

QY 1 KNEFLTNRARSRDPTFNLREVNLNFKLPPEGYILVPSTFEPNKGDCIRVFSEKKADY 60
DB 453 KNEFLTNRARSRDPTFNLREVNLNFKLPPEGYILVPSTFEPNKGDCIRVFSEKKADY 512
QY 61 QAVDEIANTANLEEDPISDDIDGFRRLFAQLAGDAEISAFELQTLIRRYLAKRODIX 119
DB 513 AELDEISADLADDEEITDDIDGFRRLFAQLAGDAEISAFELQTLIRRYLAKRODIX 572
QY 120 SDGFSIECTKIWMVMDSDSGSKLGKLEFYILMTKIOKTYREIDVDPSGTMSYEMRK 179

```

```

DB 573 TDGFSLDSCRMVNLMDSDSGSARGLVEFQLIMNKRISWLTIFKQYDLDKSGTSSYEMR 632
QY 180 KALEAGFKMPCOLHOVIYARFADDLIDFNFVRCVLRLETFKIFKOLDPENTGTIE 239
DB 633 MALESAGFKLNNKHOVIYARFADDSVDNDFVCCLVKLETFMFRFRSDPDTGTAV 692
QY 240 LDLSWLCFSV 250
DB 693 MNIAEWLLTM 703

```

RESULT 3

```

ID 08C2J1 PRELIMINARY; PRT; 713 AA.
AC 08C2J1;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Calpain 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NOD; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL: AK088547; BAC40416.1; -.
SQ SEQUENCE 713 AA; 82152 MW; 3E1FCB4D5802B864 CRC64;

```

Query Match 59.8%; Score 773.5; DB 11; Length 713;
 Best Local Similarity 56.7%; Pred. No. 1.7e-54;
 Matches 140; Conservative 58; Mismatches 48; Indels 1; Gaps 1;

```

QY 1 KNEFLTNRARSRDPTFNLREVNLNFKLPPEGYILVPSTFEPNKGDCIRVFSEKKADY 60
DB 461 KDFFLANSRAOSEFTNLREVSNRIRLPPEGYIVPSTFEPNKGDFLFFSEKKAGT 520
QY 61 QAVDEIANTANLEEDPISDDIDGFRRLFAQLAGDAEISAFELQTLIRRYLAKRODIX 119
DB 521 QELDDQIQANLPDEKVLSEEDIDNFKTLFSLKAGDDMEISVKELQTLINRISKHDLR 580
QY 120 SDGFSIECTKIWMVMDSDSGSKLGKLEFYILMTKIOKTYREIDVDPSGTMSYEMR 179
DB 581 TNGFSLESCRSVMNLMDDGNGKLGLEFNLNMRIRNYLTIFRFKFDLKGSGMSAYEMR 640
QY 180 KALEAGFKMPCOLHOVIYARFADDLIDFNFVRCVLRLETFKIFKOLDPENTGTIE 239
DB 641 MALEAGFKLNNKLEHLLITRYSEPD LAVDFNVCCLVRLTFMFRFRKLLDIDGYYT 700
QY 240 LDLSWLCFSV 250
DB 701 FDLFKWL 707

```

RESULT 4

```

ID 042133 PRELIMINARY; PRT; 715 AA.
AC 042133;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE MUC1 protein (EC 3.4.22.17).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

```

```

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Sorimachi H.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 130-715 FROM N.A.
RX MEDLINE=95260862; PubMed=7742367;
RA Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,
RA Suzuki K.;
RT "Identification of a third ubiquitous calpain species--chicken muscle
RT expresses four distinct calpains."
RL Biochim. Biophys. Acta 1261:381-393(1995).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98129049; PubMed=9467868;
RA Jeong S.-Y., Sorimachi H., Lee H.-J., Ishiura S., Suzuki K.;
RT "Molecular cloning and characterization of cDNAs for the mu-type large
RT subunit and the small subunit of chicken calpain."
RL Comp. Biochem. Physiol. 118B:539-547(1997).
DR EMBL; AB007775; BAA22659.1;
DR HSSP; P04574; 1ALV.
DR MEROPS; C02.001; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001300; Protease_C2.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF01067; Calpain_III; 1.
DR Pfam; PF00036; ehand; 3.C2; 1.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN.
DR SMART; SM00720; calpain_III; 1.
DR SMART; SM00230; Cyspc; 1.
DR SMART; SM00054; Efh; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
SQ SEQUENCE 715 AA; 81411 MW; 70679B21C5E0AD6A CRC64;

Query Match 59.2%; Score 765.5; DB 13; Length 715;
Best Local Similarity 55.1%; Pred. No. 7.8e-54;
Matches 136; Conservative 61; Mismatches 49; Indels 1; Gaps 1;

OY 1 KNFELTNRRERSDTFINLRKFLPGEYILVSTPEPNKDGFCIRVFSEKRDY 60
DB 463 RDEFLTHSSARSELFINLRKFLPGEYILVSTPEPNKDGFCIRVFSEKRDY 522
OY 61 QAVDEITANL-EEFDISEDDIDGCFRLPQLAGDAEISAFELQTLIRVLA KRDDIK 119
DB 523 EEDIDKIEAKLPDERKVSSEGEIDENFQLFRQLAGPDMELSVTELQTLIRVLA KRDDIK 582
OY 120 SDGFSIETCKIMVMDLSDSGSKGLKEFYILMTKIOKYRIYREIDVDRSGTMSYEMR 179
DB 583 TKGFSLESCRMVNLMDKDGSKGLGVEFVNLNRIYLSVFRKFDLDKSGMSATYEMR 642
OY 180 KALEEAGKMPQOLHOVIVARFADDQILIDFNFVRCVRLVETLFKIFKQOLDPENTGTIE 239
DB 643 MALPAGSKYLQKHLHLLITRYAPDLAIDDFSVCCVRLVETLFKIFKQOLDPENTGTIE 702
OY 240 LDLSISWL 246
DB 703 FGLQWL 709

RESULT 5
O918G2 PRELIMINARY; PRT; 704 AA.
AC O918G2;
DT 01-OCT-2000 (TREMBlrel. 15; Created)
DT 01-OCT-2000 (TREMBlrel. 15; Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23; Last annotation update)
DE Calpain 1 (EC 3.4.22.17).
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

```

```

OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee K.C., Evans C.W.;
RT "Zebrafish calpain 1 (capn1): Identification, expression and
RT phylogenetic implications."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF282675; AAF82808.1;
DR HSSP; P04574; 1ALV.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001300; Protease_C2.
DR InterPro; IPR000169; SHprot_acsite.
DR Pfam; PF01067; Calpain_III; 1.
DR Pfam; PF00036; ehand; 3.C2; 1.
DR Pfam; PF00648; Peptidase_C2; 1.
DR PRINTS; PR00704; CALPAIN.
DR SMART; SM00720; calpain_III; 1.
DR SMART; SM00230; Cyspc; 1.
DR SMART; SM00054; Efh; 2.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
KW Hydrolyase.
SQ SEQUENCE 704 AA; 80090 MW; E5FE0C80C18A79A8 CRC64;

Query Match 59.0%; Score 763.5; DB 13; Length 704;
Best Local Similarity 56.6%; Pred. No. 1.1e-53;
Matches 142; Conservative 51; Mismatches 57; Indels 1; Gaps 1;

OY 1 KNFELTNRRERSDTFINLRKFLPGEYILVSTPEPNKDGFCIRVFSEKRDY 60
DB 452 RDEFLTHSSARSELFINLRKFLPGEYILVSTPEPNKDGFCIRVFSEKRDY 511
OY 61 QAVDEITANL-EEFDISEDDIDGCFRLPQLAGDAEISAFELQTLIRVLA KRDDIK 119
DB 512 EEMDDKVAELPEBORDESQIDAGFKSLFRQLAGADMEISVTELQTLIRVLA KRDDIK 571
OY 120 SDGFSIETCKIMVMDLSDSGSKGLKEFYILMTKIOKYRIYREIDVDRSGTMSYEMR 179
DB 572 TDGFGKESCRMVNLMDTDSGSKGLGVEFVNLNRIYLSVFRKFDLDKSGMSATYEMR 631
OY 180 KALEEAGKMPQOLHOVIVARFADDQILIDFNFVRCVRLVETLFKIFKQOLDPENTGTIE 239
DB 632 KALEFAGKLNHLLFQILITRYEEDLSVDPDNVSCVRLVETLFKIFKSLDTPADGVIS 691
OY 240 LDLSISWLCFSV 250
DB 692 LTFEOWITLTM 702

RESULT 6
O9NOV6 PRELIMINARY; PRT; 716 AA.
AC O9NOV6;
DT 01-OCT-2000 (TREMBlrel. 15; Created)
DT 01-JUN-2002 (TREMBlrel. 21; Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23; Last annotation update)
DE Micromolar calcium activated neutral protease 1.
GN CAPN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Bovineia; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20500419; PubMed=11048924;
RA Smith T.P.L., Casas E., Rexroad C.E. III, Kappes S.M., Keele J.W.;
RT "Bovine CAPN1 maps to a region of BTA29 containing a quantitative
RT trait locus for meat tenderness."
RL J. Anim. Sci. 78:2569-2594(2000).
SQ SEQUENCE FROM N.A.

```

Query Match	58.8%	Score 761.5	DB 6	Length 716
Best Local Similarity	55.5%	Pred. No. 1.7e-53		
Matches 137	Conservative 59	Mismatches 50	Indels 1	Gaps 1
QY	1	KNFELTNARESDHFINREVLNPKLPGCYIIIVPSFEEPKKDDFCITRVSEKKADY	60	
Db	464	KDFEISNASSRAKSEDFINLREYSTRRLPGEYVVPSTFEEPKEDDFVLRFPSKSACT	523	
QY	61	QAVDEIANTL-EEFDISEDDIDDFRRLFAOLAGEDAIISAFELOTIIRRVLAKKODIK	119	
Db	524	QELDDQVQANLPDEQVLSSEEDENPKSLFRLAGDEMISSKELPTIIRIISKKKDLR	563	
QY	120	SDGSIETPCKIIVMDLDDSGSKGLKEFYIIMTKIQKQKYREIJDVDRSGTMSYEMR	179	
Db	584	TTGSELESCRSVNVNMDRNGNKKGLVPERNIIMNRIRNLSTIFRKFDLKGSGMSAYEMR	643	
QY	180	KALEPAGCKMPCOLHOVIYARFADQOLIDDFNFVACVRLTEFKIFPOLDEENGTIE	239	
Db	644	MAIEPAGCKMKKLYELITRSEPDVADVPDFNFCVCLVRLTEMFRTLIDDDGVYI	703	
QY	240	LDLISWL 246		
Db	704	FDLEFKWL 710		
RESULT 7				
Q9N185				
AC	Q9N185	PRELIMINARY	PRT	716 AA.
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)			
DE	Microtubular calcium-dependent neutral protease large subunit.			
GN	CAPN1.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Smith T.P.L., Casas E., Rexroad C.E., Kappes S.M., Keele J.W.;			
RT	"Bovine CAPN1 maps to a region containing a QTL for meat tenderness."			
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBA databases.			
DR	EMBL; AF221129; AAF32364.1;			
DR	HSSP; P04574; 1ALV.			
DR	InterPro; IPR002048; EF-hand.			
DR	InterPro; IPR001300; Protease_C2.			
DR	InterPro; IPR00169; SHprot_acsite.			
DR	Pfam; PF01067; Calpain_III; 1.			
DR	Pfam; PF00036; ehand; 3.			
DR	Pfam; PF00648; peptidase_C2; 1.			
DR	PRINTS; PR00704; CALPAIN.			
DR	SMART; SM00720; calpain_III; 1.			
DR	SMART; SM00230; Cyspc; 1.			

DR	PROSITE: PS00018; EF HAND; 2.
DR	PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
KW	Protease:
SO	SEQUENCE 716 AA; 82207 MW; A7F5C6197BD5DF23 CRC64;
	Query Match
	Best Local Similarity 58.8%; Score 761.5; DB 6; Length 716;
	Matches 137; Conservative 59; Mismatches 50; Indels 1; Gaps 1
OY	1 KNFELTRNRARRSDFTNLRREVLRNFRLPPREYLIVPSTFEFNKNKGDGCFIRFSKKADY 60
DB	464 RDEFLSNASRRSRSSQFINLREVSTFRFLRPPPELYVVPSTFEFNKGDFVLRFPSSEKSAGT 523
OY	61 QAVDEIEANT-EEFDISEDDIDGGFRRLPQLAGEADEISAFELQOTILRVLARODIK 119
DB	524 QELDQDVQANLPDQGVLSSEEIDEENFKSLFQLAGEMDIISVKELRTLLNRISHKDLR 583
OY	120 SDGSIFETCKTMVDMLDSGSGKGLKEFFYLITKIQKYOKIYREIDVDRSCTMNSYEKR 179
DB	584 TTGGSLESCRSVMNLMDRGNGKLGIVEFILMNRIINRYLSIFRKFDLDKSGSMAYEMR 643
OY	180 KALEEAGKMPCOHOVTVAVFAPDOQLIIDPNFRCIVRLLETLEKIPRODPENTGTIE 239
DB	644 MAIFEAFKLKKLKLEYLIITRTYSEPDLVADFNEVCCLVRLETMRFRKTLDTDGYYT 703
OY	240 LDLISWL 246
DB	704 FDLFKWL 710
RESULT 8	
O9GLG2	PRELIMINARY; PRT; 714 AA.
ID	O9GLG2
AC	O9GLG2;
DT	01-MAR-2001 (TREMBREL, 16, Created)
DT	01-MAR-2001 (TREMBREL, 16, Last sequence update)
DT	01-MAR-2003 (TREMBREL, 23, Last annotation update)
DE	Calpain 1.
OC	Maccaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC	Cercopithecinae; Macaca.
OX	NCBI_TaxID=9541;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE-Reclina;
RA	Nakajima T., Fukiage C., Azuma M., Shearer T.R.;
RL	"Calpain isoforms in the eye of monkey."
RL	Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
DR	EMBL: AF24440; AAC22770.1; -
DR	HSP: P04574; IALV.
DR	MEROPS: CO2.001; -
DR	InterPro: IPR0012048; EF-hand.
DR	InterPro: IPR001300; Protease_C2.
DR	InterPro: IPR000169; SHProt_acsite.
DR	Pfam: PF01067; Calpain_III; 1.
DR	Pfam: PF00036; ehand; 3.
DR	Pfam: PF00648; peptidase_C2; 1.
DR	PRINTS: PR00704; CALPAIN.
DR	SMART: SM00720; calpain_III; 1.
DR	SMART: SM00230; Cyspc; 1.
DR	SMART: SMO0054; EFh; 2.
DR	PROSITE: PS00018; EF_HAND; 2.
DR	PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
SO	SEQUENCE 714 AA; 81849 MW; 834690C214ADE7AD8 CRC64;
OY	Query Match
	Best Local Similarity 58.6%; Score 758.5; DB 6; Length 714;
	Matches 137; Conservative 57; Mismatches 52; Indels 1; Gaps 1
DB	1 KNFELTNARRRSDFTNLRREVLRNFRLPPREYLIVPSTFEFNKNKGDGCFIRFSKKADY 60
	462 RDPELANASRRSRSSQFINLREVSTFRFLPPELYVVPSTFEFNKGDFVLRFPSSEKSAGT 521

QY	6	QAVDDIEEAMN..EEDPISDDIDDDGFRRLPQALGGEAETSAEFLQTLIRLAKRQDX	113
Db	522	AEEDDDIOANLPDEOVLSDEEIDENFKALFROLAGEDEISVKELRTLIRLISKHKDC	581
QY	120	SDGFSIETCKIWMIDMDSDSGSKLGALKEFYLIWTKIQYOKIYREIDVDSRSGTMSYEHR	179
Db	582	TKFSTLSRCRSMYNLMIDRGNKKLGVENFIIMNRIRNYLSIFKPLDDSGMSAEHR	641
QY	180	KALEAGFKPCQLHOVIYVARPADDLIIDFDNFRCLVRLLETLLFKIFKOLDPENTGTE	239
Db	642	MAIESAGFKLNKKIKVELLITFRYSEPDLANVDPNFVCLVRLLETFRFRFTLIDPDLDGVA	701
QY	240	LDLISWL	246
Db	702	FDLEFKWL	708
RESULT	9		
Q8UW96			
ID	Q8UW96	PRELIMINARY;	PRT; 724 AA.
AC	Q8UW96;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, last annotation update)		
DE	Mu/m-calpain large subunit.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;		
CC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Myoshi K., Sotomachi H., Tomloka S., Ishiura S., Suzuki K.;		
RT	"Xenopus has a calpain most similar to mu/m-type chicken calpain."		
RL	Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AB061521; BAB83262.1; ..		
DR	InterPro: IPR002048; EF-hand.		
DR	InterPro: IPR001300; Protease_C2.		
DR	InterPro: IPR000169; SHprot_acsSite.		
DR	Pfam: PF01067; Calpain_III; 1.		
DR	Pfam: PF00036; efhand; 2.		
DR	Pfam: PF00648; Peptidase_C2; 1.		
DR	PRINTS: PR00704; CALPAIN.		
DR	SMART: SM00720; calpain_III; 1.		
DR	SMART: SM00230; Cyspc; 1.		
DR	SMART: SM00054; Efh; 2.		
DR	PROSITE: PS00018; EF_HAND; 2.		
DR	PROSITE: PS00139; THYOL_PROTEASE_CYS; 1.		
DR	SEQUENCE 724 AA; 81770 MW; CACCB090AE7A4288 CRC64;		

[illegible]

DB	693	LEAMEFFKALD-ESDGTAEMLNLSGEMLITMT	722
RESULT	10		
Q918T0			
ID	Q918T0	PRELIMINARY:	PRT: 702 AA.
AC	Q918T0:		
DT	01-OCT-2000	(TREMBLrel. 15, Created)	
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)	
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)	
DE	Calpain.		
GN	Cl-2.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCHI_TaxID=8335;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Cao Y., Zhao H., Grunz H.;		
RT	"A novel Xenopus gene homologous to rat calpain.";		
RL	Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AF212199; AAF63194.2; .		
DR	HSSP; P04574; IALV.		
DR	MEROPS; C02.007; .		
DR	InterPro: IPR002048; EF-hand.		
DR	InterPro: IPR001300; Protease_C2.		
DR	InterPro: IPR000169; SH3prot_acsite.		
DR	Pfam; PF01067; Calpain_IIT; 1.		
DR	Pfam; PF00036; ethand; 3.		
DR	Pfam; PF00648; Peptidase_C2; 1.		
DR	PRINTS; PR00704; CALPAIN.		
DR	SMART; SM00720; Calpain_IIT; 1.		
DR	SMART; SM00230; CysPc; 1.		
DR	PROSITE; PS00018; EF_HAND; 1.		
DR	PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.		
QO	SEQUENCE 702 AA; 79466 MW; 60666AB1D276DC51 CRC64;		

	Query Match	Similarity	51.5%	Score 667	DB 13	Length 702
	Best Local	Similarity	49.8%	Pred. No. 7.7e-46		
	Matches 125	Conservative	52	Mismatches 74	Indels 0	Gaps 0
Qy	1	KNFELTNARBRSPDIFNLREVLNFKLPDGEYLVPSTFEPKNGDCDFIRFSEKKDY	60			
Db	452	RDFLOKFTTAARSOTYIIVNREVSNRHHLPVGDYLIIVPSTFEFKNKGDCLRFSEKESKS	511			
Qy	61	QAVDDLEIANLEEFDISDEDIDDGFRRLFAOLAGEDAETISAFELQTIILRRYLAKEODIKS	120			
Db	512	LEVGDVYIAKRYEPEQISKMDVPDDEKKNLFQKLAGKEVDARELOTIILNKLSKRPDLRS	571			
Qy	121	DGFSIETCKIIVMDLSDSGSKLGKEFYIIMTKIQKYOKIYREITDVDSGTMSYEMRK	180			
Db	572	NGFTLNTFCREMISLQDDMGFTATLLSFERIILMKIQKLIATLYAKDSRSGIMSDHEJRT	631			
Qy	181	ALAEFGFMPQOLHVIYARFADDOILLDPNFVCLYRLLEFLKIFQDDPENGTIEL	240			
Db	632	ALQEGFLINKRIHESIYQRYASNDLAINFGFICAMRLLETLFRFMQPLDKSKRGVVEL	691			
Qy	241	DLISMLCFSVL	251			
Db	692	SLOERMLCATIV	702			
RESULT 11						
Q9XSJ2						
ID	Q9XSJ2	PRELIMINARY:	PRT:	709	AA.	
AC	Q9XSJ2					
DT	01-NOV-1999	(TREMBLrel. 12, Created)				
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)				
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)				
DE		Lens-specific calpain lp82.				
OS		Sus scrofa (Pig).				
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF HAND; 2.
DR PROSITE; PS00139; THIOLEPROTEASE_CYS; 1.
SQ SEQUENCE 709 AA; 82634 MW; 213150151408CC50 CRC64;

Query Match 48.8%; Score 631; DB 6; Length 709;
Best Local Similarity 46.4%; Pred. No. 6.6e-43;
Matches 116; Conservative 65; Mismatches 65; Indels 4; Gaps 2;

OY	1	KNFETNARESDPFINLRVNLNFRKLPGEYILVPSTFEPNKGDDFCIRVFSEKKADY	60
DB	454	KDFLYNASKARSKRTYIMREYSEFRLLPSEYIVPSTYEPHQDGEFILRVFSKKRLS	513
OY	61	QAVDEIEAN--LEEDISEDIDD-GFRLFAQLAGEDAEISAFELQTIILRVYLRQ	116
DB	514	EEVENTISVDRPVLEPGNTDQSESEQRQFRNIFRQIAGDMEICADELKNVLNRYVNHK	573
OY	117	DIKSDGESIEYCKIIVMDLSDSGSLGKLEFYILWTIKYOKYOKYTYREIDVDRSGTMSY	176
DB	574	DLKTGFTLESCRSMLALMDTDGSGRLNIOEFHLMKRIKTWQIKFKHYDIDQSGTINSY	633
OY	177	EMRKALEAGFPMPCQLHOVIVARFADDLITFDNFVRCIVLETLPKIFKQLDPENTG	236
DB	634	EMRNAGKDGAFHNNQLYDITMYADKYMNIDPDSFICCFVRLGEMFRANAFDPKDDG	693
OY	237	TIELDLISWL	246
DB	694	IKLNLVLEWL	703

Search completed: July 24, 2003, 12:56:54
Job time : 34.0498 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:28:44 ; Search time 6.93849 Seconds
(without alignments)
1701.191 Million cell updates/sec

Title: US-09-884-319a-6

Perfect score: 1294
Sequence: 1 KNFELTNARERSDPTFNLRL.....PENTGTLELDLSWLCPSVL 251

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1287	99.5	700	CAN2_HUMAN	P17655 homo sapien
2	1235	95.4	324	CAN2_PIG	P43367 sus scrofa
3	1232	95.2	422	CAN2_RABIT	P06814 oryctolagus
4	1225	94.7	700	CAN2_MOUSE	O08529 mus musculu
5	1210	93.5	700	CAN2_RAT	O07009 rattus norv
6	1054	81.5	700	CAN2_CHICK	O92177 gallus galli
7	799.5	61.8	705	CANX_CHICK	P00789 gallus galli
8	773.5	59.8	713	CAN1_MOUSE	O35350 mus musculu
9	772.5	59.7	713	CAN1_RAT	P97571 rattus norv
10	759.5	58.7	714	CAN1_RABIT	P06815 oryctolagus
11	757.5	58.5	714	CAN1_HUMAN	P07384 homo sapien
12	756.5	58.5	714	CAN1_PIG	P35750 sus scrofa
13	646.5	50.0	702	CANB_HUMAN	O3umg6 homo sapien
14	613	47.4	810	CAN3_HUMAN	O92177 gallus galli
15	613	47.4	821	CAN3_MOUSE	P20807 homo sapien
16	613	47.4	821	CAN3_MOUSE	O64632 mus musculu
17	609	44.6	821	CAN2_BOVIN	P16259 rattus norv
18	577	47.1	207	CAN2_BOVIN	O27971 bos taurus
19	465	35.9	263	CANS_BOVIN	P13135 bos taurus
20	465	35.9	266	CANS_PIG	P04574 sus scrofa
21	461	35.6	266	CANS_RABIT	P06813 oryctolagus
22	460	35.5	269	CANS_MOUSE	O08456 mus musculu
23	457	35.3	268	CANS_HUMAN	P04632 homo sapien
24	452	34.9	266	CANS_RAT	O64537 rattus norv
25	421	32.5	828	CAN_DROME	O11002 drosophila
26	343.5	26.5	758	CAN_SCHNA	P27730 schistosoma
27	240	18.5	198	SORC_HUMAN	P30626 homo sapien
28	236	18.2	198	SORC_CRITIO	P05044 cricetus
29	234	18.1	217	GRAN_HUMAN	P28676 homo sapien
30	204	15.8	374	CAN3_PIG	P43368 sus scrofa
31	177	13.7	171	SORC_SCHJA	O94743 schistosoma
32	173	13.4	783	YKR2_CAEEL	P34308 caenorhabdi
33	170	13.1	191	PCD6_MOUSE	P12815 mus musculu

34	168	13.0	191	PCD6_HUMAN	O75340 homo sapien
35	144.5	11.2	335	Y625_YEAST	P53238 saccharomyc
36	138.5	10.7	95	CAN1_BOVIN	O27970 bos taurus
37	126.5	9.8	169	CATR_CHIRE	P05434 chlamydomon
38	122	9.4	640	CAN5_MOUSE	O08688 mus musculu
39	116.5	9.0	147	CALM_FAGSY	O39752 tagus sylla
40	116	9.0	532	CDPK_DAUCA	P28582 daucus caro
41	116	9.0	640	CAN5_RAT	O87460 rattus norv
42	112	8.7	148	CALM_CAPAN	P93087 capiscum an
43	111.5	8.6	151	CALM_PNECA	P41041 pneumocysti
44	111.5	8.6	183	CAT3_PARTE	O27178 paramacium
45	111.5	8.6	640	CAN5_HUMAN	O15484 homo sapien

ALIGNMENTS

RESULT 1	CAN2_HUMAN	STANDARD;	PRT;	700 AA.
ID	CAN2_HUMAN	P17655; O16738; Q8WU26; Q9HBB1;		
AC	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)			
DE	(Calcium-activated neutral proteinase) (CAMP) (M-type) (M-calpain)			
DE	(Mammalian-calpain) (Calpain large polypeptide I2).			
GN	CANP2 OR CANPL2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=69166474; PubMed=2852952;			
RA	Imajoh S., Aoki K., Ohno S., Emori Y., Kawasaki H., Sugihara H.,			
RA	Suzuki K.;			
RT	"Molecular cloning of the cDNA for the large subunit of the			
RT	high-Ca2+-requiring form of human Ca2+-activated neutral protease.";			
RL	Biochemistry 27:8122-8128(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE-Astrocytoma;			
RA	MEDLINE=20403900; PubMed=10944468;			
RA	Ye Z., Connor J.R.;			
RT	"cDNA cloning by amplification of circularized first strand cDNAs			
RT	reveals non-IRE-regulated iron-responsive mRNAs.";			
RL	Biochem. Biophys. Res. Commun. 275:223-227(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE-Pancreas;			
RA	MEDLINE=22388257; PubMed=12477932;			
RA	Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Alschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Ditchenko L., Marusina K., Farmer A.F., Rubin G.W., Hong L.,			
RA	Stepleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,			
RA	Brownstein M.J., Uscin T.B., Toshimiyuki S., Carrincci F., Prange C.,			
RA	Rosa S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,			
RA	Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,			
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Whaley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,			
RA	Scherer A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[4]			

RP SEQUENCE OF 1-79 FROM N.A.
RC TISSUE-Lymph node;
RX MEDLINE=89197947; PubMed=2539381;
RA Hata A., Ohno S., Akita Y., Suzuki K.;
RT "Randomly relettered negative enhancer-like elements regulate
transcription of a human gene for the large subunit of calcium-
dependent protease.";
RL J. Biol. Chem. 264:6404-6411(1989).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=20105516; PubMed=10639123;
RA Strobl S., Fernandez-Catalan C., Braun M., Huber R., Masumoto H.,
RA Nakagawa K., Irie A., Sorimachi H., Bourenkow G., Bartunik H.,
RA Suzuki K., Bode W.;
RT "The crystal structure of calcium-free human m-calpain suggests an
electrostatic switch mechanism for activation by calcium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:588-592(2000)
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
catalyze limited proteolysis of substrates involved in
cytoskeletal remodeling and signal transduction.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
Arg-I-Xaa with Leu or Val as the P2 residue.
CC -1- COFACTOR: Binds 3 calcium ions.
CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
of calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
(regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
membrane upon Ca++ binding.
CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M23254; AAA5645.1; -;
DR EMBL: AF261089; AAP9682.1; -;
DR EMBL: BC021303; AAH21303.1; -;
DR EMBL: J04700; AAA52760.1; -;
DR PIR: S10590; C1H0H2.
DR PDB: 1KFU; 07-DEC-01.
DR PDB: 1KFX; 07-DEC-01.
DR MEROPS: C02.002; -;
DR GeneW: HGNC:1479; CAPN2.
DR MIM: 114230; -;
DR GO: GO:0008234; F:cysteine-type peptidase activity; TAS.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Protease_C2.
DR InterPro: IPR000169; SHprol_acsite.
DR Pfam: PF01067; Calpain_III; 1.
DR Pfam: PF00036; ethand; 3.
DR Pfam: PF00648; peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00720; calpain_III; 1.
DR SMART: SM00230; Cyspc; 1.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW Hydrolase; Thiol protease; Calcium-binding; Repeat; Multigene family;
KW 3D-structure; Polymorphism.
FT PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
FT CHAIN 20 700 CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
FT DOMAIN 27 355 CALPAIN.
FT DOMAIN 356 514 DOMAIN III.
FT DOMAIN 515 529 LINKER.
FT DOMAIN 530 700 DOMAIN IV.
FT CA_BIND 541 552 EF-HAND 1.
FT

FT CA_BIND 585 596 EF-HAND 2.
FT CA_BIND 615 626 EF-HAND 3.
FT DOMAIN 650 661 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT DOMAIN 680 691 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT ACT_SITE 105 105 BY SIMILARITY.
FT ACT_SITE 262 262 BY SIMILARITY.
FT ACT_SITE 286 286 BY SIMILARITY.
FT ACT_SITE 286 286 E -> D (IN dbSNP:25655).
FT VARIANT 22 22 /FTID=VAR_014435.
FT VARIANT 22 22 K -> Q (IN dbSNP:17599).
FT VARIANT 568 568 /FTID=VAR_014436.
FT CONFLICT 68 68 S -> G (IN REF. 4).
FT CONFLICT 73 74 IE -> MR (IN REF. 1).
FT CONFLICT 256 256 O -> K (IN REF. 2).
FT CONFLICT 300 300 N -> S (IN REF. 2).
FT CONFLICT 534 534 V -> F (IN REF. 3).
FT CONFLICT 4 16
FT CONFLICT 17 19
FT TURN 22 23
FT STRAND 25 26
FT HELIX 27 29
FT HELIX 32 42
FT TURN 43 43
FT TURN 43 43
FT TURN 55 58
FT TURN 69 70
FT TURN 73 76
FT STRAND 78 81
FT STRAND 86 86
FT STRAND 97 98
FT TURN 101 102
FT HELIX 105 113
FT TURN 114 115
FT HELIX 118 121
FT TURN 122 124
FT TURN 136 137
FT STRAND 138 144
FT STRAND 149 155
FT STRAND 158 161
FT TURN 162 163
FT STRAND 164 165
FT STRAND 169 170
FT STRAND 175 175
FT HELIX 177 187
FT TURN 188 189
FT HELIX 192 194
FT TURN 196 197
FT HELIX 200 206
FT TURN 207 208
FT STRAND 211 216
FT TURN 217 218
FT TURN 222 223
FT HELIX 224 224
FT TURN 233 235
FT STRAND 237 241
FT TURN 253 254
FT STRAND 264 272
FT TURN 275 276
FT STRAND 279 285
FT TURN 287 288
FT TURN 303 304
FT HELIX 310 313
FT TURN 314 317
FT STRAND 327 327
FT HELIX 328 332
FT STRAND 336 341
FT HELIX 344 346
FT TURN 351 352
FT STRAND 357 357
FT STRAND 363 365
FT TURN 367 370
FT TURN 375 376
FT TURN 378 380
FT HELIX 381 383

```

Query Match      99.5%; Score 1287; DB 1; Length 700;
Best Local Similarity 99.6%; Pred. No. 2.2e-84;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFEINLREVLNRFKLPPEGYILVPSFEPNKGDCICIRVSEKKADY 60
DB 450 KNFFLTNRARERSDFEINLREVLNRFKLPPEGYILVPSFEPNKGDCICIRVSEKKADY 509
QY 61 QAVDEIEANLEEFIDISEDIDGFRRLFAQLAGDAEISAFELQTIIRRYLAKQDIKS 120
DB 510 QAVDEIEANLEEFIDISEDIDGFRRLFAQLAGDAEISAFELQTIIRRYLAKQDIKS 569
QY 121 DGFSTETCKINWMDSDSGSKGLGKEFYILMTKIQOKYREIDVDRSGMTNMYEMRK 180
DB 570 DGFSTETCKINWMDSDSGSKGLGKEFYILMTKIQOKYREIDVDRSGMTNMYEMRK 629
QY 181 ALEAGFKMPCQLHQVIVARFADQLIIDFNFVACLRLVLETLFRIFKQDPENTGTIEL 240
DB 630 ALEAGFKMPCQLHQVIVARFADQLIIDFNFVACLRLVLETLFRIFKQDPENTGTIEL 689
QY 241 DLISWLCFSVL 251
DB 690 DLISWLCFSVL 700

RESULT 2
CAN2_PIG STANDARD; PRT; 324 AA.
AC P43367;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calpain 2, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated
DE neutral proteinase) (CAMP) (M-type) (M-calpain) (Millimolar-calpain)
DE (Fragment).
GN CAPN2.
OS Sus scrofa (Pig).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN [1]
RP SEQUENCE OF 1-209 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=94146155; PubMed=8312396;
RA Sun W., Ji S.O., Ebert P.J., Bidwell C.A., Hancock D.L.;
RT "Cloning the partial cDNAs of mu-calpain and m-calpain from porcine
RT skeletal muscle."
RL Biochimie 75:931-936(1993).
RN [2]
RP SEQUENCE OF 122-324 FROM N.A.
RC TISSUE=pulmonary artery;
RX MEDLINE=9840921; PubMed=9728040;
RA Zhang J.L., Patel J.M., Block E.R.;
RT "Hypoxia-specific upregulation of calpain activity and gene
RT expression in pulmonary artery endothelial cells."
RL Am. J. Physiol. 275:1461-1468(1998).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction (by similarity).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
CC Arg-I-Xaa with Leu or Val as the p2 residue.
CC -1- COFACTOR: Binds 3 calcium ions.
CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
CC of calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding.
CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC This SWISS-Prot entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: U01181; AAC48401.1; -.
DB EMBL: U71320; AAB17381.1; -.
DB HSSP: Q07009; IDFO.
DB MEROPS: C02.002; -.
DB InterPro: IPR002048; EF-hand.
DB InterPro: IPR001300; Protease_C2.
DB InterPro: IPR00169; SHprot_acsite.
DB Pfam: PF01067; Calpain_III; 1.
DB Pfam: PF00036; ethand; 2.
DB PRINTS: PR00704; CALPAIN.
DB SMART: SM00720; calpain_III; 1.
DB SMART: SM00054; EFh; 2.
DB PROSITE: PS00018; EF_HAND; 1.
DB PROSITE: PS00139; THIOL_PROTEASE_CYS; PARTIAL.
DB PROSITE: PS00639; THIOL_PROTEASE_HIS; PARTIAL.
DB PROSITE: PS00640; THIOL_PROTEASE_ASN; PARTIAL.
KW Hydrolase; Thiol protease; Calcium-binding; Repeat; Multigene family.
FT NON_TER 1 1
FT DOMAIN <1 157 III.
FT DOMAIN 158 324 IV, CALCIUM-BINDING.
FT DOMAIN <1 138 DOMAIN III.
FT DOMAIN 139 153 LINKER.
FT DOMAIN 154 324 DOMAIN IV.
FT CA_BIND 165 176 EF_HAND 1.
FT CA_BIND 209 220 EF_HAND 2.
FT CA_BIND 239 250 EF_HAND 3.
FT DOMAIN 274 285 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT DOMAIN 304 315 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT CONFLICT 202 202 R -> K (IN REF. 2).
SQ SEQUENCE 324 AA; 37808 MW; 3929553239E123CF CRC64;

Query Match      95.4%; Score 1235; DB 1; Length 324;
Best Local Similarity 95.6%; Pred. No. 4.6e-81;
Matches 240; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFEINLREVLNRFKLPPEGYILVPSFEPNKGDCICIRVSEKKADY 60
DB 74 KNFFLTNRARERSDFEINLREVLNRFKLPPEGYILVPSFEPNKGDCICIRVSEKKADY 133
QY 61 QAVDEIEANLEEFIDISEDIDGFRRLFAQLAGDAEISAFELQTIIRRYLAKQDIKS 120
DB 134 QAVDEIEANLEEFIDISEDIDGFRRLFAQLAGDAEISAFELQTIIRRYLAKQDIKS 193
QY 121 DGFSTETCKINWMDSDSGSKGLGKEFYILMTKIQOKYREIDVDRSGMTNMYEMRK 180
DB 194 DGFSTETCKINWMDSDSGSKGLGKEFYILMTKIQOKYREIDVDRSGMTNMYEMRK 253
QY 181 ALEAGFKMPCQLHQVIVARFADQLIIDFNFVACLRLVLETLFRIFKQDPENTGTIEL 240
DB 254 ALEAGFKMPCQLHQVIVARFADQLIIDFNFVACLRLVLETLFRIFKQDPENTGTIEL 313
QY 241 DLISWLCFSVL 251
DB 314 DLISWLCFSVL 324

RESULT 3
CAN2_RABIT STANDARD; PRT; 422 AA.
AC P06814;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calpain 2, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated
DE neutral proteinase) (CAMP) (M-type) (M-calpain) (Millimolar-calpain)
DE (Fragment).

```

GN CAPN2.
 OS Oryctolagus cuniculus (Rabbit).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86250902; PubMed=2424911;
 RA Emori Y., Kawasaki H., Sugihara H., Imajoh S., Kawashima S.,
 RA Suzuki K.;
 RT "Isolation and sequence analyses of cDNA clones for the large
 RT subunits of two isozymes of rabbit calcium-dependent protease";
 RL J. Biol. Chem. 261:9463-9471(1986).
 RN [2]
 RP CALCIUM-BINDING DATA.
 RX MEDLINE=87279982; PubMed=3038855;
 RA Minami Y., Emori Y., Kawasaki H., Suzuki K.;
 RT "E-F hand structure-domain of calcium-activated neutral protease
 RT (CAMP) can bind Ca²⁺ ions";
 RL J. Biochem. 101:889-895(1987).
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodelling and signal transduction (By similarity).
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the p2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions.
 CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
 CC of calcium and inhibited by calpastatin.
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca²⁺ binding.
 CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M13797; AA31455.1; -;
 DR PIR: B24815; B24815.
 DR HSSP: Q07009; IDP0.
 DR MEROPS: C02.002; -;
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001300; Protease_C2.
 DR InterPro: IPR000169; SHProt_acsite.
 DR Pfam: PF01067; Calpain_III; 1.
 DR Pfam: PF00036; ehand; 3.
 DR Pfam: PF00648; peptidase_C2; 1.
 DR PRINTS: PRO0704; CALPAIN.
 DR SMART: SM00720; calpain_III; 1.
 DR SMART: SM00054; EFh_3.
 DR PROSITE: PS00018; EF_HAND; 2.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; PARTIAL.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; PARTIAL.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; PARTIAL.
 KW Hydrolyase; Thiol protease; Calcium-binding; Multigene family.
 FT NON_TER 1 1
 FT DOMAIN <1 77 CALPAIN.
 FT DOMAIN 78 236 DOMAIN III.
 FT DOMAIN 237 251 LINKER.
 FT DOMAIN 252 422 DOMAIN IV.
 FT CA_BIND 263 274 EF-HAND 1.
 FT CA_BIND 307 318 EF-HAND 2.
 FT CA_BIND 337 348 EF-HAND 3.
 FT DOMAIN 372 383 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT DOMAIN 402 413 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
 FT ACT_SITE 8 8 BY SIMILARITY.
 SQ SEQUENCE 422 AA: 49494 MW: AFAFA3C48A333C41 CRG64;

Query Match 95.2%; Score 1232; DB 1; Length 422;
 Best Local Similarity 93.6%; Pred. No. 1e-80;
 Matches 235; Conservative 11; Mismatches 5; Indels 0; Gaps 0;
 OY 1 KNFFLTNRARRSPTFTNLREVLNRFKLPPEEYLLVSTFPPNNDGDFCIIVFSEKKADY 60
 DB 172 KNFFLTNRARRSPTFTNLREVLNRFKLPPEEYLLVSTFPPNNDGDFCVAFSEKKADY 231
 OY 61 QAVDEIFANLEEDPDISDDDDGFRRLFAQAGEDAEISAFELQTLIRVLARODIKS 120
 DB 232 QAVDEIFADLEADVSEDDDDGFRRLFAQAGEDAEISAFELQTLIRVLARODIKT 291
 OY 121 DGFSEITCKINVDMLDSGSGKLGLKEFYILMTKIQKYQKYREIVDRSGTMSYEMRK 180
 DB 292 DGLSEITCKINVDMLDSGSGTKLGLKEFYVLMFKYQKYREIVDRSGTMSYEMRK 351
 OY 181 ALEAGFRMPQQLNOVIYARFADQQLITDPNFRCVRLFTLTKIKQIDPENTGTEL 240
 DB 352 ALEAGFRMPQQLHEVIYARFADQQLITDPNFRCVRLFTLTKIKQIDPNTGMQL 411
 OY 241 DLISMICPSVL 251
 DB 412 DLISMICPSVL 422
 RESULT 4
 CAN2_MOUSE STANDARD: PRT; 700 AA.
 ID CAN2_MOUSE 008529; 035518; 054843;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)
 DE (Calcium-activated neutral protease) (CAMP) (M-type) (M-calpain)
 DE (Molecular-calpain) (80 kDa M-calpain subunit) (CAL80).
 GN CAPN2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/c; PubMed=9339374;
 RX MEDLINE=97480729; PubMed=9339374;
 RA Dear T.N., Matena K., Vingron M., Boehm T.;
 RT "A new subfamily of vertebrate calpains lacking a calmodulin-like
 RT domain: implications for calpain regulation and evolution";
 RL Genomics 45:175-184(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/c;
 RA Ozaki Y.;
 RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-CNS;
 RA Glass J.D., Nash N.R., Dry I., Culver D., Wesselingh S.;
 RT "Cloning of m-calpain from mouse nervous system";
 RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyze limited proteolysis of substrates involved in
 CC cytoskeletal remodelling and signal transduction (By similarity).
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the p2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions.
 CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
 CC of calcium and inhibited by calpastatin.
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca²⁺ binding.
 CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.

```

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Y10139; CAJ1227.1; -
DR EMBL: D38117; BAA22964.1; -
DR EMBL: AF015038; AAB94029.1; -
DR HSP: Q07009; IDP0.
DR MEROPS: C02.002; -.
DR MGD: MGI:88264; Capn2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Protease_C2.
DR InterPro: IPR000169; SHProl_acsite.
DR Pfam: PF01067; Calpain_III; 1.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN_III; 1.
DR SMART: SM00720; calpain_III; 1.
DR SMART: SM00230; Cyspc; 1.
DR SMART: SM00054; Eph; 2.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW Hydrolyase; Thiol protease; Calcium-binding; Repeat; Multigene family.
FT PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).
FT CHAIN 20 700 CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.
FT DOMAIN 20 355 CALPAIN.
FT DOMAIN 356 514 CALPAIN III.
FT DOMAIN 515 529 LINKER.
FT DOMAIN 530 700 DOMAIN IV.
FT CA_BIND 541 552 EF-HAND 1.
FT CA_BIND 585 596 EF-HAND 2.
FT CA_BIND 615 626 EF-HAND 3.
FT DOMAIN 650 661 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT DOMAIN 680 691 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT ACT_SITE 105 105 BY SIMILARITY.
FT ACT_SITE 262 262 BY SIMILARITY.
FT ACT_SITE 286 286 BY SIMILARITY.
FT CONFLICT 194 194 A -> T (IN REF. 1).
FT CONFLICT 212 212 A -> G (IN REF. 2).
FT CONFLICT 402 402 E -> G (IN REF. 1).
SQ SEQUENCE 700 AA; 79871 MW; 6821465290968316 CRC64;
Query Match 94.7%; Score 1225; DB 1; Length 700;
Best Local Similarity 93.6%; Pred. No. 5.5e-80;
Matches 235; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

```

```

RESULT 5
ID CAN2_RAT STANDARD; PRT; 700 AA.
AC 007009;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)
DE (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)
DE (Mollinolar-calpain).
GN CAPN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94032492; PubMed=8218419;
RA Deluca C.I., Davies P.L., Samis J.A., Elce J.S.;
RT "Molecular cloning and bacterial expression of cDNA for rat calpain
RT II 80 kDa subunit.";
RL Biochim. Biophys. Acta 1216:81-93(1993).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=21240297; PubMed=11342050;
RA Moldoveanu T., Hosfield C.M., Jia Z., Elce J.S., Davies P.L.;
RT "Ca(2+)-induced structural changes in rat m-calpain revealed by
RT partial proteolysis.";
RL Biochim. Biophys. Acta 1545:245-254(2001).
RN [3]
RP MUTAGENESIS OF LYS-230; LYS-234 AND GLU-504.
RX MEDLINE=21269273; PubMed=11102442;
RA Hosfield C.M., Moldoveanu T., Davies P.L., Elce J.S., Jia Z.;
RT "Calpain mutants with increased Ca2+ sensitivity and implications for
RT the role of the C(2)-like domain.";
RL J. Biol. Chem. 276:7404-7407(2001).
RN [4]
RP MUTAGENESIS OF CYS-105; HIS-262; ASN-286 AND TRP-288.
RX MEDLINE=95361909; PubMed=7635186;
RA Arthur J.S., Gauthier S., Elce J.S.;
RT "Active site residues in m-calpain: identification by site-directed
RT mutagenesis.";
RL FEBS Lett. 368:397-400(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=20069318; PubMed=10601010;
RA Hosfield C.M., Elce J.S., Davies P.L., Jia Z.;
RT "Crystal structure of calpain reveals the structural basis for
RT Ca(2+)-dependent protease activity and a novel mode of enzyme
RT activation.";
RL EMBO J. 18:6880-6889(1999).
RN [6]
RP FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
RP catalyze limited proteolysis of substrates involved in
RP cytoskeletal remodelling and signal transduction.
RN [7]
RP CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
RP Arg-I-Xaa with Ieu or Val as the P2 residue.
RN [8]
RP COFACTOR: Binds 3 calcium ions.
RN [9]
RP ENZYME REGULATION: Activated by 200-1000 micromolar concentrations
RP of calcium and inhibited by calpastatin.
RN [10]
RP SUBUNIT: Heterodimer of a large (catalytic) and a small
RP (regulatory) subunit.
RN [11]
RP SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
RP membrane upon Ca++ binding.
RN [12]
RP SIMILARITY: Contains 5 EF-hand calcium-binding domains.
RN [13]
RP SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL, 109120; AAA16327.1; -

DR PIR: S38361; S38361.

DR PDB: 1DFO; 26-NOV-01.

DR MEROPS: C02.002; -

DR InterPro: IPR002048; EF-hand.

DR InterPro: IPR001300; Protease_C2.

DR InterPro: IPR000169; SHprol.acsite.

DR Pfam: PF01067; Calpain_III; 1.

DR Pfam: PF00036; ehand; 3.

DR Pfam: PF00648; peptidase_C2; 1.

DR PRINTS: PR00704; CALPAIN.

DR SMART: SM00230; calpain_III; 1.

DR SMART: SM00054; CysPc; 1.

DR PROSITE: PS00018; EF-HAND; 2.

DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.

DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.

DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.

KW Hydrolase: Thiol protease; Calcium-binding; Repeat; Multigene family; 3D-structure.

KW PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).

FT CHAIN 20 700 CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.

FT DOMAIN 20 355 CALPAIN.

FT DOMAIN 356 514 DOMAIN III.

FT DOMAIN 515 529 LINKER.

FT DOMAIN 530 700 DOMAIN IV.

FT CA_BIND 541 552 EF-HAND 1.

FT CA_BIND 585 596 EF-HAND 2.

FT CA_BIND 615 626 EF-HAND 3.

FT DOMAIN 650 661 ANCESTRAL CALCIUM SITE 4.

FT DOMAIN 680 691 ANCESTRAL CALCIUM SITE 5.

FT ACT_SITE 105 105

FT ACT_SITE 262 262

FT ACT_SITE 286 286

FT MUTAGEN 105 105 C->S: DECREASE OF ACTIVITY.

FT MUTAGEN 105 105 K->S: DECREASE OF 12% OF THE ACTIVITY.

FT MUTAGEN 226 226 K->S: NO EFFECT.

FT MUTAGEN 230 230 K->E: DECREASE OF 84% OF THE ACTIVITY.

FT MUTAGEN 234 234 K->E: DECREASE OF 85% OF THE ACTIVITY.

FT MUTAGEN 234 234 K->S: DECREASE OF 20% OF THE ACTIVITY.

FT MUTAGEN 262 262 H->A: LOSS OF ACTIVITY.

FT MUTAGEN 286 286 N->A: LOSS OF ACTIVITY.

FT MUTAGEN 288 288 W->Y: DECREASE OF 95% OF THE ACTIVITY.

FT MUTAGEN 504 504 E->S: DECREASE OF 10% OF THE ACTIVITY.

SO SEQUENCE 700 AA; 79919 MW; 296B0DC3BEEF5B90 CRC64;

Query Match 93.5%; Score 1210; DB 1; Length 700;

Best Local Similarity 92.4%; Pred. No. 6; 4e-79;

Matches 232; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERDFFINLRVLRNFKLPPEGYIIVPSPFEENKGDGFCIRVFSEKKADY 60

DB 450 KNFFLTNRARERDFFINLRVLRNFKLPPEGYIIVPSPFEENKGDGFCIRVFSEKKADY 509

QY 61 QAVDEIEANLEEDFISDDIDGFRRLFAOLAGDAEIASAELQTLIRVLAKRODKS 120

DB 510 QVDEIEANLEEDFISDDIDGFRRLFAOLAGDAEIASAELQTLIRVLAKRODKS 569

QY 121 DGFSEIETCKIMVMDLSDSGSLGKLEFIITMTKIOKOKIYREIDVDRSGTMSYEMRK 180

DB 570 DGFSEIETCKIMVMDLSDSGSLGKLEFIITMTKIOKOKIYREIDVDRSGTMSYEMRK 629

QY 181 ALEEGFKMPCOLHOVIYARFADPOLIIDDFNVFVCLVLETLFIIFKOLDPENGTIEL 240

DB 630 ALEEGFKMPCOLHOVIYARFADPOLIIDDFNVFVCLVLETLFIIFKOLDPENGTIEL 689

QY 241 DLISWLCFSVL 251

DB 690 DLISWLCFSVL 700

RESULT 6

CC CAN2_CHICK STANDARD; PRT; 700 AA.

ID 092178; 01-NOV-1997 (Rel. 35, Created)

AC 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Calpain 2, large [catalytic] subunit precursor (EC 3.4.22.17)

DE (Calcium-activated neutral proteinase) (CANP) (M-type) (M-calpain)

DE (Molecular-calpain).

GN CANP2.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OK NCBI_TaxId=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Muscle;

RC MEDLINE=95260862; PubMed=7742367;

RA Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S., Suzuki K.

RA "Identification of a third ubiquitous calpain species -- chicken muscle expresses four distinct calpains."

RL Biochim. Biophys. Acta 1261:381-393(1995).

CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which catalyze limited proteolysis of substrates involved in cytoskeletal remodeling and signal transduction (By similarity).

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or Arg-I-Xaa with Ileu or Val as the P2 residue.

CC -1- COFACTOR: Binds 3 calcium ions.

CC -1- ENZYME REGULATION: Activated by 200-1000 micromolar concentrations of calcium and inhibited by calpastatin.

CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small (regulatory) subunit.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma membrane upon Ca++ binding.

CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL, D38026; BA07228.1; -

DR PIR: S57194; S57194.

DR HSSP: Q07009; 1DFO.

DR MEROPS: C02.002; -

DR InterPro: IPR002048; EF-hand.

DR InterPro: IPR001300; Protease_C2.

DR InterPro: IPR000169; SHprol.acsite.

DR Pfam: PF01067; Calpain_III; 1.

DR Pfam: PF00036; ehand; 3.

DR Pfam: PF00648; peptidase_C2; 1.

DR PRINTS: PR00704; CALPAIN.

DR SMART: SM00230; calpain_III; 1.

DR SMART: SM00230; CysPc; 1.

DR SMART: SM00054; Efh; 3.

DR PROSITE: PS00018; EF-HAND; 2.

DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.

DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.

DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.

KW Hydrolase: Thiol protease; Calcium-binding; Multigene family.

KW PROPEP 1 19 ANCHORS TO THE SMALL SUBUNIT (POTENTIAL).

FT CHAIN 20 700 CALPAIN 2, LARGE [CATALYTIC] SUBUNIT.

FT DOMAIN 20 355 CALPAIN.

FT DOMAIN 356 514 DOMAIN III.

FT DOMAIN 515 529 LINKER.

```

FT DOMAIN 530 700 DOMAIN IV.
FT CA_BIND 541 552 EF-HAND 1.
FT CA_BIND 541 552 EF-HAND 2.
FT CA_BIND 541 552 EF-HAND 3.
FT CA_BIND 541 552 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT CA_BIND 541 552 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT ACT_SITE 105 105 BY SIMILARITY.
FT ACT_SITE 262 262 BY SIMILARITY.
FT ACT_SITE 286 286 BY SIMILARITY.
SQ SEQUENCE 700 AA; 79228 MW; C3AEDB39CCB56D3B CRC64;

Query Match 81.5%; Score 1054; DB 1; Length 700;
Best Local Similarity 78.5%; Pred. No. 8e-68;
Matches 197; Conservative 32; Mismatches 22; Indels 0; Gaps 0;

QY 1 KNEFLTRARERSDFINIREVLNRRKLPGEYILVPSFEPNKGDFCIRVFSEKKADY 60
DB 450 KNEFLTRARERSDFINIREVLNRRKLPAGEYILVPSFEPNKGDFCIRVFSEKKANS 509
QY 61 QAVDEIEANLEEFIDSEDDIDGFRRLPAOLAGEAETSAPELOTILRLVLAARDIKS 120
DB 510 TVIDDEIEANFEETEDDEDIEPSKKLFGOLAGDAETSAPELOTILRLVLAARDIKS 569
QY 121 DEFSTETCKINVDMLSDSGKLGKEFYILMTKIOKYOIYREIDVDKSGTMNSYEMRK 180
DB 570 DEFSTETCKINVDMLSDSGKLGKEFYILMTKIOKYOIYREIDVDKSGTMNSYEMRK 629
QY 181 ALEBAGFKPKCOLHOYIVARFADDLIDPDPNFRVRLVLELEFKFOLDENGTIEL 240
DB 630 ALEBAGFKPKCOLHOYIVARFADDELIDPDPNFRVRLVLELEFKFOLDENGTIEL 689
QY 241 DLISWLCFSVL 251
DB 690 NLIWMLFETVI 700

RESULT 7
CANX_CHICK STANDARD; PRT; 705 AA.
ID CANX_CHICK P00789;
AC 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calpain, large [catalytic] subunit (EC 3.4.22.17) (Calcium-
DE activated neutral proteinase) (CANP) (Mu/M-type).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061606; PubMed=6095110;
RA Ohno S., Emori Y., Imaizumi S., Kawasaki H., Kisaragi M., Suzuki K.;
RT "Evolutionary origin of a calcium-dependent proteinase by fusion of
RT genes for a thiol protease and a calcium-binding protein";
RL Nature 312:566-570(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86082358; PubMed=3000828;
RA Emori Y., Ohno S., Tobita M., Suzuki K.;
RT "Gene structure of calcium-dependent proteinase retains the ancestral
RT organization of the calcium-binding protein gene.";
RL FEBS Lett. 194:249-252(1986).
RN [3]
RP CALCIUM-BINDING DATA.
RX MEDLINE=87279982; PubMed=3038855;
RA Minami Y., Emori Y., Kawasaki H., Suzuki K.;
RT "E-F hand structure-domain of calcium-activated neutral proteinase
RT (CANP) can bind Ca2+ ions.";
RL J. Biochem. 101:889-895(1987).
RN [4]
RP CHARACTERIZATION.

```

```

RX MEDLINE=95260862; PubMed=7742367;
RA Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,
RA Suzuki K.;
RT "Identification of a third ubiquitous calpain species -- chicken
RT muscle expresses four distinct calpains.";
RL Biochim. Biophys. Acta 1261:381-393(1995).
RT FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyzes limited proteolysis of substrates involved in
CC cytoskeletal remodeling and signal transduction.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-1-Xaa, Met-1-Xaa or
CC Arg-1-Xaa with Leu or Val as the P2 residue.
CC -1- COFACTOR: Binds 3 calcium ions.
CC -1- ENZYME REGULATION: Activated by micromolar concentrations of
CC calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Heterodimer of large (catalytic) and a small (regulatory)
CC subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding (By similarity).
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC -1- CAUTION: THIS PROTEIN WAS PREVIOUSLY THOUGHT TO BE M-CALPAIN BUT
CC HAS SINCE BEEN FOUND TO BE AN INTERMEDIATE FORM BETWEEN THE M AND
CC MU TYPES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X01415; CA25658.1; -
DR PIR: A00979; CICH.
DR HSP: Q07009; IDFO.
DR MEROPS: C02.003; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Protease_C2.
DR InterPro: IPR000169; SHprot_acstie.
DR Pfam: PF01067; Calpain_III; 1.
DR Pfam: PF00036; efhand; 2.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00720; calpain_III; 1.
DR SMART: SM00230; Cyspc; 1.
DR SMART: SM00054; Eph; 2.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
KW Hydrolyase; Thiol protease; Calcium-binding; Multigene family.
FT DOMAIN 23 358 CALPAIN.
FT DOMAIN 359 517 DOMAIN III.
FT DOMAIN 518 533 LINKER.
FT DOMAIN 534 704 DOMAIN IV.
FT CA_BIND 545 556 EF-HAND 1.
FT CA_BIND 545 556 EF-HAND 2.
FT CA_BIND 545 556 EF-HAND 3.
FT CA_BIND 545 556 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT CA_BIND 545 556 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT ACT_SITE 108 108 BY SIMILARITY.
FT ACT_SITE 265 265 BY SIMILARITY.
FT ACT_SITE 289 289 BY SIMILARITY.
SQ SEQUENCE 705 AA; 80351 MW; ABCDDC56298E8AA CRC64;

Query Match 61.8%; Score 799.5; DB 1; Length 705;
Best Local Similarity 58.2%; Pred. No. 1e-49;
Matches 146; Conservative 58; Mismatches 46; Indels 1; Gaps 1;

QY 1 KNEFLTRARERSDFINIREVLNRRKLPGEYILVPSFEPNKGDFCIRVFSEKKADY 60

```


Db 453 KDEFLRNSRANSEFFINLREVSNOIRLPGEYIVPSTFEPBKADFLRVEFKOSDT 512
 QY 61 QAVDEIEANL-EEFDISEDDIDDFRRLFAQAGEAIEISAFELQITILIRYLAKRODIK 119
 Db 513 AELDEIEISADLDEDEIEDDIEDGFKNNFQOLAGEDMISVFEKLTILNRIYARKDK 572
 QY 120 SDGSEIEFCKIWMMLSDSGSKGLKEFYIIWTKIQYKQYKIREIDVRSCTMNSYEMR 179
 Db 573 TDFGSLDSCRMVNMMDKSGARLGLVEFQILMNRIRSLVTFRQYDLDSKSGTMSSEMR 632
 QY 180 KALEAGFRMPCOLHOVIVARFADQOLIIPDNFRCIVRLTETLKIFKQOLDEPNTGTE 239
 Db 633 MALEAGFRKLNKLHQLVAVARADAEVDFNFCVCLVLETFMFRFHSMDRGITAV 692
 QY 240 IDLISMLCFSV 250
 Db 693 MFLAEWLLTLM 703
 RESULT 8
 CAN1_MOUSE STANDARD: PRT: 713 AA.
 AC 035350: 088666;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 41, Last annotation update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated neutral proteinase) (CAMP) (Mu-type) (mucANP) (Micromolar-calpain).
 GN CAPN1 OR CAPN1 OR CAPN1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxId=10090;
 RX [1]
 RC SEQUENCE FROM N.A.
 RA Polier C., Poussard S., Faust D.M., Imaizumi-Sherrier T., Weiss M.C., Ducasting A., Montarras D., Panset C., Guenet J.-U.;
 RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129;
 RA Sahr K.E., Andrabi S., Peters L.L., Chishti A.H.;
 RT "Cloning and characterization of the cDNA and gene encoding the mouse mu-calpain large subunit protein."
 RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
 CC -i- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which catalyze limited proteolysis of substrates involved in cytoskeletal remodelling and signal transduction.
 CC -i- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or Arg-I-Xaa with Leu or Val as the P2 residue.
 CC -i- COFACTOR: Binds 3 calcium ions (By similarity).
 CC -i- ENZYME REGULATION: Activated by micromolar concentrations of calcium and inhibited by calpastatin.
 CC -i- SUBUNIT: Heterodimer of a large (catalytic) and a small (regulatory) subunit.
 CC -i- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma membrane upon Ca++ binding (By similarity).
 CC -i- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -i- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF021847; AAB72222.1; -
 DR EMBL: AF084459; AAC33134.1; -
 DR HSSP: A07009; IDFO.
 DR MEROPS: C02.001; -.

DR MGI:88263; Capn1.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001300; Protease_C2.
 DR InterPro: IPR000169; Shprot_acsite.
 DR Pfam: PF00036; Calpain_III; 1.
 DR Pfam: PF00036; etHand; 3.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00720; calpain_III; 1.
 DR SMART: SM00230; calpain_III; 1.
 DR SMART: SM00054; EFh; 3.
 DR PROSITE: PS00018; EF-HAND; 2.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
 KW Hydroxylase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 30 365
 FT DOMAIN 366 525
 FT DOMAIN 526 541
 FT DOMAIN 542 712
 FT CA_BIND 553 564
 FT CA_BIND 597 608
 FT CA_BIND 627 638
 FT DOMAIN 652 673
 FT DOMAIN 692 703
 FT ACT_SITE 115 115
 FT ACT_SITE 272 272
 FT ACT_SITE 296 296
 FT CONFLICT 488 488
 FT CONFLICT 696 696
 SQ SEQUENCE 713 AA: 82106 MW; 3E1E26C95802B864 CRC64;
 Query Match 59.8%; Score 773.5; DB 1; Length 713;
 Best Local Similarity 56.7%; Pred. No. 7.2e-48;
 Matches 140; Conservative 58; Mismatches 48; Indels 1; Gaps 1;
 QY 1 KNEFITNARESDFFINLREVLNFKRLPGEYIVPSTFEPBKADFLRVEFKADY 60
 Db 461 RQFLNANRSGSEHFFINLREVSNNRILPGEYIVPSTFEPBKADFLRFSKKG 520
 QY 61 QAVDEIEANL-EEFDISEDDIDDFRRLFAQAGEAIEISAFELQITILIRYLAKRODIK 119
 Db 521 QELDQIQANLPDENLSEEDIDNFKILFSLADDMEISVKELOITILNRIASHKRLR 580
 QY 120 SDGSEIEFCKIWMMLSDSGSKGLKEFYIIWTKIQYKQYKIREIDVRSCTMNSYEMR 179
 Db 581 TNGFELESRSVNMMDKSGARLGLVEFQILMNRIRSLVTFRQYDLDSKSGTMSSEMR 640
 QY 180 KALEAGFRMPCOLHOVIVARFADQOLIIPDNFRCIVRLTETLKIFKQOLDEPNTGTE 239
 Db 641 MALEAGFRKLNKLHQLVAVARADAEVDFNFCVCLVLETFMFRFHSMDRGITAV 700
 QY 240 IDLISML 246
 Db 701 FDLFKWL 707
 RESULT 9
 CAN1_RAT STANDARD: PRT: 713 AA.
 AC P97571;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated neutral proteinase) (CAMP) (Mu-type) (mucANP) (Micromolar-calpain).
 GN CAPN1 OR CCSI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxId=10116;
 RX [1]
 RP SEQUENCE FROM N.A.


```

RX MEDLINE-97107433; PubMed-8950173;
RA Sohmaichi H., Amano S., Ishiura S., Suzuki K.;
RT "Primary sequences of rat mu-calpain large and small subunits are,
RL Biochim. Biophys. Acta 1309:37-41(1996).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
CC Arg-|-Xaa with Leu or Val as the P2 residue.
CC -1- COFACTOR: Binds 3 calcium ions (by similarity).
CC -1- ENZYME REGULATION: Activated by micromolar concentrations of
CC calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding (By similarity).
CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U53858; AAC53001.1; -.
DR PDB: 1KXR; 2O-MAR-02.
DR MEROPS: C02.001; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Protease_C2.
DR InterPro: IPR000169; SHprot_acsite.
DR Pfam: PF01067; Calpain_III; 1.
DR Pfam: PF00036; ehand; 3.
DR Pfam: PF00648; peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00720; calpain_III; 1.
DR SMART: SM00230; Cyspc; 1.
DR SMART: SM00054; EFh; 3.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE NEG.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE NEG.
KM Calcium-binding; Hydrolyase; Thiol protease; 3D-structure.
FT DOMAIN 30 365 CALPAIN.
FT DOMAIN 366 525 DOMAIN III.
FT DOMAIN 526 541 LINKER.
FT DOMAIN 542 712 DOMAIN IV.
FT CA_BIND 553 564 EF_HAND 1.
FT CA_BIND 597 608 EF_HAND 2.
FT CA_BIND 627 638 EF_HAND 3.
FT DOMAIN 662 673 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
FT DOMAIN 692 703 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT ACT_SITE 115 115 BY SIMILARITY.
FT ACT_SITE 272 272 BY SIMILARITY.
FT ACT_SITE 296 296 BY SIMILARITY.
SQ SEQUENCE 713 AA; 82118 MW; 6E66460B0EFEBB CRC64;

Query Match 59.7%; Score 772.5; DB 1; Length 713;
Best Local Similarity 56.7%; Pred. No. 8.4e-48;
Matches 140; Conservative 58; Mismatches 48; Indels 1; Gaps 1;

```

```

DB 581 TNGFSLSCRSVMNIMDRDGNKGLGVEFNILMNIRVYLTFRRKFDLCKSGSMAYEMR 640
OY 180 KALENAGFRMPCQHQVYVARADQQLIIDPNFRCVLRLETLFKIRKQDPEWTGITE 239
DB 641 MAIEAAGRLNKKLHETITRYSEPDLAVDENFVCCVLRLEMTMRPFKILDTDLGGVT 700
OY 240 IDLISML 246
DB 701 FDLEFKML 707

RESULT 10
ID CAN1_RABIT STANDARD; PRT; 302 AA.
AC P06815;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated
DE neutral proteinase) (CAMP) (mu-type) (mucAMP) (Micromolar-calpain)
DE (Fragment).
GN CAPN1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-86250902; PubMed-2424911;
RA Emori Y., Kawasaki H., Sugihara H., Imajob S., Kawashima S.,
RA Suzuki K.;
RT "Isolation and sequence analyses of cDNA clones for the large
RT subunits of two isoforms of rabbit calcium-dependent protease."
RL J. Biol. Chem. 261:9465-9471(1986).
[2]
RN CALCINUM-BINDING DATA.
RX MEDLINE-87279982; PubMed-3038855;
RA Minami Y., Emori Y., Kawasaki H., Suzuki K.;
RT "E-F hand structure-domain of calcium-activated neutral protease
RT (CAMP) can bind Ca2+ ions."
RL J. Biochem. 101:889-895(1987).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodelling and signal transduction.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Met-|-Xaa or
CC Arg-|-Xaa with Leu or Val as the P2 residue.
CC -1- COFACTOR: Binds 3 calcium ions (by similarity).
CC -1- ENZYME REGULATION: Activated by micromolar concentrations of
CC calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding (By similarity).
CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M1363; AA31456.1; -.
DR PIR: A24815; A24815.
DR HSSP: 007009; IDEO.
DR MEROPS: C02.001; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Protease_C2.
DR InterPro: IPR000169; SHprot_acsite.
DR Pfam: PF01067; Calpain_III; 1.

```

DR Pfam: PF00036; ehand: 3.
 DR SMART: SM00720; calpain_III: 1.
 DR SMART: SM00054; Eph: 3.
 DR PROSITE: PS00018; EF_HAND: 2.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS: PARTIAL.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS: PARTIAL.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN: PARTIAL.
 DR Hydrolyase: Thiol protease; Calcium-binding; Multigene family.
 FT NON_TER 1
 FT DOMAIN 1
 FT LINKER 114
 FT DOMAIN III 115
 FT LINKER 130
 FT DOMAIN IV 131
 FT CA_BIND 142
 FT CA_BIND 153
 FT CA_BIND 186
 FT CA_BIND 197
 FT CA_BIND 216
 FT CA_BIND 227
 FT CA_BIND 251
 FT CA_BIND 262
 FT CA_BIND 281
 FT CA_BIND 292
 FT CA_BIND 302
 FT CA_BIND 35274
 FT CA_BIND 35274 MM; 178BFEF4216CEAB CRC64;
 SO SEQUENCE 302 AA; 35274 MM; 178BFEF4216CEAB CRC64;
 Query Match 58.7%; Score 759.5; DB 1; Length 302;
 Best Local Similarity 55.9%; Pred. No. 2.7e-47;
 Matches 138; Conservative 56; Mismatches 52; Indels 1; Gaps 1;
 OY 1 KNFFLTNARSRSDFFINLRVFNFKLPGEYILVPTSPFPNKGDCIRFESKADY 60
 DB 50 KDFELANASRAKSEDFINLRVSTFRKLPGEYVVPSTFEPNKGDFVLRFSKRG 109
 OY 61 QAVDEILANL-EERDISSEDIDGFRRLFAQLAGEAIEISAFELQTLIRVLAKRODIK 119
 DB 110 QELDQIQANLPDEQVLAELAEIDENFKALFRLAGEDELEISVRELQTLIRNRTSHKDLR 169
 OY 120 SDGSIETCKIMVMDLSDGSGKGLKEFYILMTKIQYKTYRRIIDVDRSTNNSYEMR 179
 DB 170 TKGSEMSCRSMVNLMDRGNGKGLVFEFNLMTNRYLAIFRFEDIDSGSMAYEMR 229
 OY 180 KALEFAGKMPOLHOVIVARFADPOLITDPNFCVRLETLEKIFKOLDPEMNTGTE 239
 DB 230 MAIESAGKINKKLYELITRISPEPLAVDPNFCVLCVLRLETMRFRKTLDTLDGVT 289
 OY 240 LDLISWL 246
 DB 290 FDLFKWL 296
 RESULT 11
 CAN1_HUMAN STANDARD; PRT; 714 AA.
 ID CAN1_HUMAN STANDARD; PRT; 714 AA.
 AC P07384;
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated
 DE neutral protease) (CaMP) (Mu-type) (mucAMP) (Micro-molar-calpain).
 GN CAPN1 OR CAPN1L.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RX MEDLINE=66301172; PubMed=3017764;
 RX Aoki K., Imajo S., Ohno S., Emori Y., Koike M., Kosaki G., Suzuki K.,
 RX "Complete amino acid sequence of the large subunit of the low-Ca2+-
 RX requiring form of human Ca2+-activated neutral protease (mucAMP)
 RX deduced from its cDNA sequence."
 RX FEBS Lett. 205:313-317(1986).
 RP SEQUENCE FROM N.A.
 RP MEDLINE=90380278; PubMed=2400579;
 RP Sorimachi H., Ohmi S., Emori Y., Kawasaki H., Saido T.C., Ohno S.,
 RP Minami Y., Suzuki K.,
 RP "A novel member of the calcium-dependent cysteine protease family."
 RP Biol. Chem. Hoppe-Seyler 371:171-176(1990).

RM [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strassberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Tashyuk S., Carinelli P., Prange C.,
 RA Raha S.S., Loguella N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McKernan P.J., McKernan K.J., Males K.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyik S.W.,
 RA Villano D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shvachenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyzes limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
 CC Arg-I-Xaa with Leu or Val as the p2 residue.
 CC -1- COFACTOR: Binds 3 calcium ions.
 CC -1- ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin.
 CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
 CC membrane upon Ca++ binding (By similarity).
 CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X04366; CAA27881.1; -.
 CC EMBL: BC008751; AAH08751.1; -.
 CC PIR: A26213; CIRH.
 CC HSP: Q07009; IDFO.
 CC MEROPS: C02.001; -.
 CC GeneW: HGNC:1476; CAPN1.
 CC MIM: 114220; -.
 CC GO: GO:0008284; P:positive regulation of cell proliferation; TAS.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001300; Protease_C2.
 DR InterPro: IPR000169; SHprot_acsite.
 DR Pfam: PF01067; Calpain_III; 1.
 DR Pfam: PF00036; ehand: 3.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00720; calpain_III; 1.
 DR SMART: SM00230; Cyspc; 1.
 DR SMART: SM00054; Eph; 3.
 DR PROSITE: PS00018; EF_HAND; 2.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
 KW Hydrolyase: Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 30
 FT DOMAIN 365
 FT DOMAIN 366
 FT DOMAIN 527
 FT DOMAIN 542
 FT LINKER.

```

FT DOMAIN 543 713 DOMAIN IV.
FT CA_BIND 554 565 EF-HAND 1.
FT CA_BIND 598 609 EF-HAND 2.
FT CA_BIND 628 639 EF-HAND 3.
FT DOMAIN 663 674 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT DOMAIN 693 704 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT ACT_SITE 115 115 By similarity.
FT ACT_SITE 272 272 By similarity.
FT ACT_SITE 296 296 By similarity.
FT CONFLICT 548 548 K -> N (IN REF. 3).
SQ SEQUENCE 714 AA: 81889 MW: 1CB6D/C56D063498 CRC64:

Query Match 58.5%; Score 757.5; DB 1; Length 714;
Best Local Similarity 55.5%; Pred. No. 9,9e-47;
Matches 137; Conservative 57; Mismatches 52; Indels 1; Gaps 1;

QY 1 KNFELTRARERSDFINLREVNLNFKLPGEYIIVPSFEPNKGDFCIRVSEKKADY 60
DB 462 RDFFLANASRRASEQFINREVSTFRRLPGEYVVPSTFEPNKGDFVLRFSSEKAGT 521
QY 61 QAVDEIEANTL-EEFDISEDDIDGFRRLFAQLAGEDAIEISAFELQTLIRVLAARODIK 119
DB 522 VELDDQIDANLPDEQVLSSEEDENFKALFRLAGEDMEISVRELRTILNRIISKHDLR 581
QY 120 SDGFSIETCKIWMIDSDSGSKGLKEFYIIMTKYQYKIREIDYDRSGTNSYEMR 179
DB 582 TKGFSLESCRMVNMIDRDNGKIGLVEFNILMNIRNYLSIFRKFEDLDKSGMSAYEMR 641
QY 180 KALEBAGKMPCCQLHQVYVARPADQLIIDEDNVEYRCVRETLFKIRKDDPENTGITE 239
DB 642 MAIESAGFKLNKKILYELITRYSEPLAVDFDNFVCCVLRULEMTREFKTLDTDIDGVYT 701
QY 240 LDIISML 246
DB 702 FDLERKWL 708

RESULT 12
CAN1_PIG STANDARD; PRT; 714 AA.
AC P35750: Q29600; Q9N0M6;
DT 01-JUN-1994 (Rel. 29; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Calpain 1, large [catalytic] subunit (EC 3.4.22.17) (Calcium-activated
  neutral proteinase) (CANP) (Mu-type) (mucanp) (Microtubular-calpain).
GN CANP1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith T.P.L., Simmen F.A., Vallet J.A.;
RT "Rapid Communication: nucleotide sequences of two isoforms of porcine
  microtubular calcium-activated neutral proteinase 1 (mucalpain) cDNA.";
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE OF 326-415 FROM N.A.
RC TISSUE=Small intestine;
RA Winfree A.K., Fredholm M., Davies W.;
RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
[3]
RP SEQUENCE OF 528-623 FROM N.A.
RC TISSUE=Skeletal muscle;
RA MEDLINE=94146155; PubMed=8312396;
RA Sun W., Ji S.O., Ebert P.J., Bidwell C.A., Hancock D.L.;
RT "Cloning the partial cDNAs of mu-calpain and m-calpain from porcine
  skeletal muscle.";
RL Biochimie 75:931-936(1993).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
  catalyze limited proteolysis of substrates involved in
  cytoskeletal remodelling and signal transduction.

```

```

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
  Arg-I-Xaa with Leu or Val as the P2 residue.
CC -1- COFACTOR: Binds 3 calcium ions (By similarity).
CC -1- ENZYME REGULATION: Activated by micromolar concentrations of
  calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
  (regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
  membrane upon Ca++ binding (By similarity).
CC -1- SIMILARITY: Contains 5 EF-hand calcium-binding domains.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF263610; AAF73444.1; -
DR EMBL: F14611; CA23154.1; -
DR EMBL: U01180; AAA65125.1; -
DR HSPD: Q07009; IDPO.
DR MEROPS: C02.001; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001300; Protease_C2.
DR InterPro: IPR000169; SHprot_Lacsite.
DR Pfam: PF01067; Calpain_I1; 1.
DR Pfam: PF00036; efhand; 3.
DR Pfam: PF00648; Peptidase_C2; 1.
DR PRINTS: PR00704; CALPAIN.
DR SMART: SM00720; Calpain_I1; 1.
DR SMART: SM00230; Cyspc; 1.
DR SMART: SM00054; Efh; 3.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE NEG.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE NEG.
KM Hydroxylase, Thiol protease; Calcium-binding; Multigene family.
FT DOMAIN 30 365 CALPAIN.
FT FT 366 526 DOMAIN I.
FT FT 527 542 LINKER.
FT FT 543 713 DOMAIN IV.
FT FT 544 565 EF-HAND 1.
FT FT 558 588 EF-HAND 2.
FT FT 598 609 EF-HAND 3.
FT FT 628 639 EF-HAND 4.
FT FT 663 674 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT FT 693 704 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
FT FT 704 714 BY SIMILARITY.
FT FT 115 115 BY SIMILARITY.
FT FT 272 272 BY SIMILARITY.
FT FT 296 296 BY SIMILARITY.
FT FT 328 528 V -> I (IN REF. 3).
FT FT 531 531 I -> N (IN REF. 3).
FT FT 541 541 S -> G (IN REF. 3).
FT FT 622 622 S -> A (IN REF. 3).
SQ SEQUENCE 714 AA: 81738 MW: 0BB31DB4FC56363A CRC64:

Query Match 58.5%; Score 756.5; DB 1; Length 714;
Best Local Similarity 55.5%; Pred. No. 1,2e-46;
Matches 137; Conservative 57; Mismatches 52; Indels 1; Gaps 1;

QY 1 KNFELTRARERSDFINLREVNLNFKLPGEYIIVPSFEPNKGDFCIRVSEKKADY 60
DB 462 RDFFLANASRRASEQFINREVSTFRRLPGEYVVPSTFEPNKGDFVLRFSSEKAGT 521
QY 61 QAVDEIEANTL-EEFDISEDDIDGFRRLFAQLAGEDAIEISAFELQTLIRVLAARODIK 119
DB 522 VELDDQVQALPDEQVLSSEEDENFKALFRLAGEDMEISVRELRTILNRIISKHDLR 581
QY 120 SDGFSIETCKIWMIDSDSGSKGLKEFYIIMTKYQYKIREIDYDRSGTNSYEMR 179
DB 582 TKGFSLESCRMVNMIDRDNGKIGLVEFNILMNIRNYLSIFRKFEDLDKSGMSAYEMR 641

```

QY 180 KALEBAGFKMPCOLHOVIVARFADDDLIIDDFNVRCLVRELETKIFKOLDPENTGTIE 239
 DB 642 MAIESAGFKLNKKLFELLITRYSERDLAVEDNFVCCIVRLTETMRFRFKLTDLTGVT 701
 QY 240 LDLISWL 246
 DB 702 FDLFRWL 708

RESULT 13

CANB_HUMAN STANDARD: PRT: 702 AA.
 AC Q9UMQ6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Calpain 11 (EC 3.4.22.17) (Calcium-activated neutral proteinase 11)
 DE (CAMP 11).
 GN CAPN11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9933989; PubMed=10409436;
 RA Dear T.N., Moller A., Boehm T.;
 RT "CAPN11: A calpain with high mRNA levels in testis and located on
 RT chromosome 6.";
 RL Genomics 59:243-247(1999).
 CC -1 FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
 CC catalyzes limited proteolysis of substrates involved in
 CC cytoskeletal remodeling and signal transduction.
 CC -1 CATALYTIC ACTIVITY: Preferential cleavage: Tyr-Xaa, Met-Xaa or
 CC Arg-Xaa with Leu or Val as the P2 residue.
 CC -1 SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -1 TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS.
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
 CC -1 SIMILARITY: Contains 4 EF-hand calcium-binding domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AJ242832; CAB52137.1; -
 DR HSSP: Q07009; IDP0.
 DR MEROPS: C02.013; -
 DR Gene: HGNC:1478; CAPN11.
 DR MIM: 604822; -
 DR GO: GO:0004198; F:calpain activity; TAS.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001300; Protease_C2.
 DR InterPro: IPR000169; SHprot_acsite.
 DR Pfam: PF01067; Calpain_III.1.
 DR Pfam: PF00036; ehand; 2.1.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00720; calpain_III; 1.
 DR SMART: SM00230; Cyspc; 1.
 DR SMART: SM00054; Efh; 2.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
 KW Hydrolase; Thiol protease; Calcium-binding; Multigene family.
 FT DOMAIN 17
 FT DOMAIN 353 514 DOMAIN III.

FT DOMAIN 515 530 LINKER.
 FT DOMAIN 531 701 DOMAIN IV.
 FT CA_BIND 586 597 EF_HAND 1.
 FT CA_BIND 616 627 EF_HAND 2.
 FT DOMAIN 651 662 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT DOMAIN 681 692 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT ACT_SITE 102 102 BY SIMILARITY.
 FT ACT_SITE 259 259 BY SIMILARITY.
 FT ACT_SITE 283 283 BY SIMILARITY.
 SQ SEQUENCE 702 AA; 80582 MW; 3A618682B9674FAE CRC64;

Query Match 50.0%; Score 646.5; DB 1; Length 702;
 Best Local Similarity 49.4%; Pred. No. 7.6e-39;
 Matches 122; Conservative 50; Mismatches 74; Indels 1; Gaps 1;

QY 1 KNFELTNRRSRSTFINLRVLRNFKLPPEYILVSTFEPNKGDFCINVFSEKRAY 60
 DB 450 KEFTKYDDHGFSEFTNSREVSQRLPRPEYIIIPSTFEPHNDADFLRVFEKISES 509
 QY 61 QAVDD-EIEANLEEFEDISEDIDGFRRLFAQLAGEDAEISAEFLQTLRRVLAKRODIK 119
 DB 510 WELDEVNVAEDLOEBKVEDMDQDLHLFKIVAGEKEIGVYELQRLNMAIKRFSK 569
 QY 120 SDGSIETCKIMVMDLSDSGSKIGLKEFYLLMKIQYQKTYREIDVNRSGTANSTEMR 179
 DB 570 TKGGLDRCRCIMINLMDKSGKGLGLEFKILMKRKKWMDIFRECDQDHSGTNSYEMR 629
 QY 180 KALEBAGFKMPCOLHOVIVARFADDDLIIDDFNVRCLVRELETKIFKOLDPENTGTIE 239
 DB 630 LVIERAGIKLNKKVQVIVARFADDDLIIDDFNVRCLVRELETKIFKOLDPENTGTIE 239
 QY 240 LDLISWL 246
 DB 690 LSLBQWL 696

RESULT 14

CAN3_CHICK STANDARD: PRT: 810 AA.
 ID CAN3_CHICK 092177;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Calpain 3 large subunit (EC 3.4.22.17) (Calpain I3) (Calpain p94,
 DE large [catalytic] subunit) (Calcium-activated neutral proteinase 3)
 DE (CAMP 3) (Muscle-specific calcium-activated neutral protease 3 large
 DE subunit).
 GN CAPN3.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=95260862; PubMed=7742367;
 RA Sorimachi H., Tsukahara T., Okada-Ban M., Sugita H., Ishiura S.,
 RA Suzuki K.;
 RT "Identification of a third ubiquitous calpain species -- chicken
 RT muscle expresses four distinct calpains.";
 RL Biochim. Biophys. Acta 1261:381-393(1995).
 CC -1 FUNCTION: Calcium-regulated non-lysosomal thiol-protease.
 CC -1 CATALYTIC ACTIVITY: Preferential cleavage: Tyr-Xaa, Met-Xaa or
 CC Arg-Xaa with Leu or Val as the P2 residue.
 CC -1 ENZYME REGULATION: Activated by micromolar concentrations of
 CC calcium and inhibited by calpastatin.
 CC -1 SUBUNIT: Heterodimer of a large (catalytic) and a small
 CC (regulatory) subunit.
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic and nuclear (potential).
 CC -1 TISSUE SPECIFICITY: SKELETAL MUSCLE. LOW LEVELS IN SPLEEN,
 CC -1 INTERESTING AND BONE.
 CC -1 SIMILARITY: Contains 4 EF-hand calcium-binding domains.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D38028; BAA07230.1; -
 DR PIR: S57196; S57196.
 DR HSSP: Q07009; 1DF0.
 DR MEROPS: C02.004; -
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001300; Protease_C2.
 DR InterPro: IPR000169; SHproL_acsite.
 DR Pfam: PF01067; Calpain_III; 1.
 DR Pfam: PF00036; efhand; 4.
 DR Pfam: PF00648; Peptidase_C2; 1.
 DR PRINTS: PR00704; CALPAIN.
 DR SMART: SM00720; calpain_III; 1.
 DR SMART: SM00230; Cyspc; 1.
 DR SMART: SM00054; Eph; 3.
 DR PROSITE: PS00018; EF_HAND; 2.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; FALSE_NEG.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; FALSE_NEG.
 KW Hydrolase; Thiol protease; Calcium-binding; Nuclear protein;
 KW Multigene family.
 FT DOMAIN 43 421 CALPAIN.
 FT DOMAIN 422 579 DOMAIN III.
 FT DOMAIN 580 638 LINKER.
 FT DOMAIN 639 809 DOMAIN IV.
 FT CA_BIND 694 705 EF-HAND 1 (PROBABLE).
 FT CA_BIND 724 735 EF-HAND 2 (PROBABLE).
 FT DOMAIN 759 770 ANCESTRAL CALCIUM SITE 3 (POTENTIAL).
 FT DOMAIN 789 800 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
 FT ACT_SITE 123 123 BY SIMILARITY.
 FT ACT_SITE 327 327 BY SIMILARITY.
 FT ACT_SITE 351 351 BY SIMILARITY.
 SQ SEQUENCE 810 AA; 93560 MW; E8DE941IC51041E CRC64;
 Query Match 47.4%; Score 613; DB 1; Length 810;
 Best local similarity 39.3%; Pred. No. 2, le-36;
 Matches 114; Conservative 65; Mismatches 67; Indels 44; Gaps 2;

DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Calpain 3 large subunit (EC 3.4.22.17) (calpain L3) (calpain p94,
 DE large [catalytic] subunit) (Calcium-activated neutral proteinase 3)
 DE (CAMP 3) (muscle-specific calcium-activated neutral protease 3 large
 DE subunit) (incl.1).
 GN CAPN3 OR CAMP3 OR CAMP3L3 OR NCCL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANTS LGMD2A.
 RX MEDLINE=95236448; PubMed=7720071;
 RA Richard I., Brox O., Allamand V., Fougerousse F.,
 RA Chianmikulchai N., Bourg N., Brenguier L., Devaud C.,
 RA Pasturaud P., Roudaut C., Hillaire D., Passos-Bueno M.-R., Zatz M.,
 RA Tischfield J.A., Fardeau M., Jackson C.E., Cohen D., Beckmann J.S.;
 RT "Mutations in the proteolytic enzyme calpain 3 cause limb-girdle
 RT muscular dystrophy type 2A.";
 RL Cell 81:27-40(1995).
 RN [2]
 RP SEQUENCE OF 44-821 FROM N.A.
 RX MEDLINE=90062125; PubMed=2555341;
 RA Sorimachi H., Imaizumi S., Emori Y., Kawasaki H., Ohno S.,
 RA Minami Y., Suzuki K.;
 RT "Molecular cloning of a novel mammalian calcium-dependent protease
 RT distinct from both m- and mu-types. Specific expression of the mRNA
 RT in skeletal muscle.";
 RL J. Biol. Chem. 264:20106-20111(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS II AND III).
 RA Dickson J.M.J., Love D., Evans C.W.E.;
 RT "Alternatively exon-spliced isoforms of calpain 3 expressed in human
 RT leukocytes.";
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=99264234; PubMed=10330340;
 RA Richard I., Roudaut C., Saenz A., Pogue R., Grinbergen J.E.M.A.,
 RA Anderson L.V.B., Beley C., Cobo A.-W., de Diego C., Eymard B.,
 RA Gallano P., Ginjaar H.B., Jasa A., Pollitt C., Topaloglu H.,
 RA Urtizberea J.A., de Visser M., van der Kooij A., Bushby K., Bakker E.,
 RA Lopez de Munain A., Fardeau M., Beckmann J.S.;
 RT "Calpainopathy -- a survey of mutations and polymorphisms.";
 RL Am. J. Hum. Genet. 64:1524-1540(1999).
 RN [5]
 RP VARIANTS LGMD2A GLN-572 AND GLY-744.
 RX MEDLINE=96176567; PubMed=8624690;
 RA Fardeau M., Hillaire D., Mignard C., Feingold N., Feingold J.,
 RA Mignard D., de Ubéda B., Collin H., Tome F.M.S., Richard I.,
 RA Beckmann J.S.;
 RT "Juvenile limb-girdle muscular dystrophy. Clinical, histopathological
 RT and genetic data from a small community living in the Reunion
 RT island.";
 RL Brain 119:295-308(1996).
 RN [6]
 RP VARIANTS LGMD2A.
 RX MEDLINE=97294404; PubMed=9150160;
 RA Richard I., Brenguier L., Dincer P., Roudaut C., Bady B.,
 RA Burgunder J.-M., Chemaly R., Garcia C.A., Halaby G., Jackson C.E.,
 RA Kurut D.M., Lefranc G., Legum C., Loiselet J., Merlin L.,
 RA Nivelet-Duchevalier A., Ollagnon-Roman E., Restagno G., Topaloglu H.,
 RA Beckmann J.S.;
 RT "Multiple independent molecular etiology for limb-girdle muscular
 RT dystrophy type 2A patients from various geographical origins.";
 RL Am. J. Hum. Genet. 60:1128-1138(1997).
 RN [7]
 RP VARIANTS LGMD2A ASN-336; GLN-490; VAL-702 AND GLN-748.
 RX MEDLINE=97410018; PubMed=9266733;
 RA Dincer P., Leturcq F., Richard I., Piccolo F., Yalnizoglu D.,
 RA de Toma C., Akcoeren Z., Brox O., Deburgrave N., Brenguier L.,
 RA Roudaut C., Urtizberea J.A., Jung D., Tan E., Jeanpierre M.,

CC RA Campbell K.F., Kaplan J.-C., Beckmann J.S., Topaloglu H.;
CC RA A biochemical, genetic, and clinical survey of autosomal recessive
RT limb girdle muscular dystrophies in Turkey.";
RL Ann. Neurol. 42:222-229(1997).
RN [8]
RP VARIANTS LGMD2A ARG-222; GLU-486; TRP-489 AND GLN-748.
RX MEDLINE=98433925; PubMed=9762961;
RA Urtasun M., Saenz A., Roudaut C., Posa J.J., Urticbera J.A.,
RA Cobo A.-M., Richard I., Garcia Bragado F., Letourneau F., Kaplan J.-C.,
RA Marti Masso J.F., Beckmann J.S., Lopez de Munain A.,
RT "Limb-girdle muscular dystrophy in Guipuzcoa (Basque Country,
RT Spain)";
RL Brain 121:1735-1747(1998).
RN [9]
RP VARIANT LGMD2A 200-PHE--LEU-204 DEL.
RX MEDLINE=98112493; PubMed=9452114;
RA Haefliger K., Speer A., Huebner C., Voit T., Oexle K.;
RT "A small in-frame deletion within the protease domain of muscle-
RT specific calpain, p94 causes early-onset limb-girdle muscular
RT dystrophy 2A.";
RL Hum. Mutat. Suppl. 1:S298-S300(1998).
RN [10]
RP VARIANT LGMD2A GLY-744.
RX MEDLINE=98317835; PubMed=9655129;
RA Penissou-Beslier I., Richard I., Dubas F., Beckmann J.S., Fardeau M.;
RT "pseudometabolic expression and phenotypic variability of calpain
RT deficiency in two siblings.";
RL Muscle Nerve 21:1078-1080(1998).
RN [11]
RP VARIANT LGMD2A CYS-360.
RX MEDLINE=98442618; PubMed=9771675;
RA Kawai H., Akaike M., Kunishige M., Inui T., Adachi K., Kimura C.,
RA Kawajiri M., Nishida Y., Endo I., Kashiwagi S., Nishino H.,
RA Fujiwara T., Okuno S., Roudaut C., Richard I., Beckmann J.S.,
RA Miyoshi K., Matsumoto T.;
RT "Clinical, pathological, and genetic features of limb-girdle muscular
RT dystrophy type 2A with new calpain 3 gene mutations in seven patients
RT from three Japanese families";
RL Muscle Nerve 21:1493-1501(1998).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Met-I-Xaa or
CC Arg-I-Xaa with Leu or Val as the P2 residue.
CC -1- ENZYME REGULATION: Activated by micromolar concentrations of
CC calcium and inhibited by calpastatin.
CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=I;
CC IsoId=P20807-1; Sequence=Displayed;
CC Name=II;
CC IsoId=P20807-2; Sequence=VSP_005227, VSP_005228;
CC Name=III;
CC IsoId=P20807-3; Sequence=VSP_005229;
CC -1- TISSUE SPECIFICITY: Skeletal muscle.
CC -1- DISEASE: Defects in CAPN3 are the cause of limb-girdle muscular
CC dystrophy type 2A (LGMD2A) [MIM:253601]. LGMD2A is both autosomal
CC dominantly and recessively transmitted. It is characterized by
CC progressive symmetrical atrophy and weakness of the proximal limb
CC muscles and elevated serum creatine kinase. The symptoms usually
CC begin during the first two decades of life, and the disease
CC gradually worsens, often resulting in loss of walking ability 10
CC or 20 years after onset.
CC -1- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C2.
CC -1- DATABASE: NAME=Leiden Muscular Dystrophy pages;
CC NOTE=Calpain-3 mutations in LGMD2A;
CC WWW="http://www.dmd.nl/capn3_home.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```
CC      use by non-profit institutions as long as its content is in no way  
CC      modified and this statement is not removed. Usage by other parties  
CC      entails requires a license agreement (See http://www.isb-sdb.cn/announce/  
CC      or send an email to licenselsb@sib.ch).  
DR      EMBL; X85030; CAA59403.1; -  
DR      EMBL; AF127764; AAD28253.1; -  
DR      EMBL; AF127765; AAD28254.3; -  
DR      PIR; A56218; CIRH03.  
DR      HSSP; Q07009; IDPO.  
DR      MEROPS; CO2_004; -.  
DR      GeneW; HGNC:1480; CAPN3.  
DR      MIM; 114240; -.  
DR      MIM; 253600; -.  
DR      GO; GO:0005622; C:intracellular; TAS.  
DR      GO; GO:0004198; F:calpain activity; TAS.  
DR      GO; GO:0008234; F:cysteine-type peptidase activity; TAS.  
DR      GO; GO:0004871; F:signal transducer activity; TAS.  
DR      GO; GO:0007517; P:muscle development; TAS.  
DR      InterPro; IPR002048; EF-hand.  
DR      InterPro; IPR001300; Protease_C2.  
DR      InterPro; IPR00169; SHprol_acsite.  
DR      Pfam; PF01067; Calpain_I1; 1.  
DR      Pfam; PF00036; ehfand; 3.  
DR      Pfam; PF00648; Peptidase_C2; 1.  
DR      PRINTS; PR00704; CALPAIN.  
DR      SMART; SMO0720; calpain_i1i; 1.  
DR      SMART; SMO0230; Cyspc; 1.  
DR      SMART; SMO0054; Eph; 3.  
DR      PROSITE; PS00018; EF-HAND; 2.  
DR      PROSITE; PS00139; THIO_L_PROTEASE_CYS; 1.  
DR      PROSITE; PS00639; THIO_L_PROTEASE_HIS; FALSE_NEG.  
DR      PROSITE; PS00640; THIO_L_PROTEASE_ASN; FALSE_NEG.  
KV      Hydrolase; Thiol protease; Calcium-binding; Multigene family; Repeat;  
KM      Disease mutation; Polymorphism; Alternative splicing.  
FT      DOMAIN          49       428  
FT                      CAIPAIN.  
FT      DOMAIN          429       586  
FT                      LINKER.  
FT      DOMAIN          587       649  
FT                      LINKER.  
FT      DOMAIN          650       820  
FT                      DOMAIN_IV.  
FT      CA_BIND         705       716  
FT                      EF-HAND 1 (PROBABLE).  
FT      CA_BIND         735       741  
FT                      EF-HAND 2 (PROBABLE).  
FT      DOMAIN          770       786  
FT                      ANCESTRAL CALCIUM SITE 3 (POTENTIAL).  
FT      DOMAIN          800       811  
FT                      ANCESTRAL CALCIUM SITE 4 (POTENTIAL).  
FT      ACT_SITE        129       129  
FT                      BY SIMILARITY.  
FT      ACT_SITE        334       334  
FT                      BY SIMILARITY.  
FT      ACT_SITE        358       358  
FT                      BY SIMILARITY.  
FT      VARSPLIC        268       315  
FT                      /FTID=VSP_005227.  
FT      VARSPLIC        595       639  
FT                      KKKKTKPIIFVSDDRNKNKLGVDOSEBGKGKTSPDKOKO  
FT                      SPDP -> P (In isoform II).  
FT                      /FTID=VSP_005228.  
  
Query Match               47.4%; Score 613; DB 1; Length 821;  
Best Local Similarity    39.3%; Pred. No. 2; 2e-36;  
Matches 116; Conservative 66; Mismatches 63; Indels 50; Gaps 3
```

DB 761 QLYDITMRYADKHMNIDFDSFICFVRLEGMRFAFHAFDKDDGIIKLNLEWL 815

Search completed: July 24, 2003, 12:52:27
Job time : 8.93849 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:46:04 ; Search time 12.3158 Seconds
(without alignments)
1959.945 Million cell updates/sec

Title: US-09-884-319a-6
Perfect score: 1294
Sequence: 1 KNFFLTNRARERSDTFINLR.....PENTGTIEDLISWLCFSL 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 segs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1287	99.5	700	1 C1RH02	calpain (EC 3.4.22
2	1232	95.2	422	2 B24815	calpain (EC 3.4.22
3	1210	93.5	700	2 S38361	calpain (EC 3.4.22
4	1054	81.8	700	2 S57194	calpain (EC 3.4.22
5	799.5	61.8	705	1 C1CHH	calpain (EC 3.4.22
6	759.5	58.7	320	2 A24815	calpain (EC 3.4.22
7	757.5	58.5	714	1 C1RH0	calpain (EC 3.4.22
8	739.5	57.1	586	2 S57195	calpain (EC 3.4.22
9	633.5	49.0	703	2 A48764	calpain (EC 3.4.22
10	613	47.4	810	1 S57196	calpain (EC 3.4.22
11	613	47.4	821	1 C1RH03	calpain (EC 3.4.22
12	609	47.1	821	1 B34488	calpain (EC 3.4.22
13	465	35.9	263	2 A34466	calpain (EC 3.4.22
14	465	35.9	263	1 C1PGL	calpain (EC 3.4.22
15	461	35.6	266	1 C1RBL	calpain (EC 3.4.22
16	457	35.3	268	1 C1RH0L	calpain (EC 3.4.22
17	442.5	34.2	184	2 A55143	calpain (EC 3.4.22
18	421	32.5	805	1 A55054	calpain (EC 3.4.22
19	338.5	26.2	758	1 A39343	calpain (EC 3.4.22
20	240	18.5	198	2 S52094	calpain (EC 3.4.22
21	236	18.2	198	2 A25706	calpain (EC 3.4.22
22	234	18.1	217	2 A42578	calpain (EC 3.4.22
23	173	13.4	653	1 D44749	calpain (EC 3.4.22
24	168.5	13.0	186	2 D84673	calpain (EC 3.4.22
25	164.5	12.7	153	2 S04970	calpain (EC 3.4.22
26	144.5	11.2	335	2 S64352	calpain (EC 3.4.22
27	126.5	9.8	169	1 BCKM	calpain (EC 3.4.22
28	118.5	9.2	541	2 F96776	calpain (EC 3.4.22
29	116	9.0	425	2 S17759	calpain (EC 3.4.22

30	116	9.0	532	2 T14335	protein kinase, ca
31	114.5	8.8	520	2 F85059	probable calcium d
32	111.5	8.6	151	2 JN0722	calmodulin - Pneum
33	111.5	8.6	183	2 S71319	centrin ICLC - Pa
34	111	8.6	149	2 S22860	calmodulin 2 (clon
35	111	8.6	162	2 JN0066	slow cardiac tropo
36	110.5	8.5	150	2 H86194	hypothetical prote
37	110	8.5	317	1 A31797	Spec-related prote
38	109	8.4	149	1 S5306	calmodulin - leaf
39	109	8.4	149	1 S22503	calmodulin [simila
40	109	8.4	149	2 H84667	calmodulin (cam2)
41	109	8.4	187	2 T02887	probable calmoduli
42	109	8.4	465	1 T03024	calcium-dependent
43	109	8.4	508	1 A43713	calcium-dependent
44	109	8.4	647	2 JCS772	tissue-specific ca
45	108.5	8.4	157	2 D96689	calmodulin-related

ALIGNMENTS

RESULT 1

C1RH02
calpain (EC 3.4.22.17) large chain 2 - human
N/Alternate names: calpain chain L-2; calpain II catalytic chain; high-calcium regul
C/Species: Homo sapiens (man)
C/Date: 21-Nov-1993 #sequence, revision 09-Aug-1997 #text, change 16-Jul-1999
C/Accession: S10590; A31218; A33529
R/Sort: MacInt, H.; Ohn, S.; Emori, Y.; Kawasaki, H.; Saido, T.C.; Ohno, S.; Minami, Y.
Biol. Chem. Hoppe-Seyler 371(Suppl.), 171-176, 1990
A/Title: A novel member of the calcium-dependent cysteine protease family.
A/Reference number: S10589; MIM:60380278; PMID:2400579
A/Accession: S10590
A/Molecule type: mRNA
A/Residues: 1-700 <SOR>
R/Imajo, S.; Aoki, K.; Ohno, S.; Emori, Y.; Kawasaki, H.; Sugihara, H.; Suzuki, K.
Biochemistry 27, 8122-8128, 1988
A/Title: Molecular cloning of the cDNA for the large subunit of the high-Ca(2+)-requi
A/Reference number: A31218; MIM:89166474; PMID:2852952
A/Accession: A31218
A/Molecule type: protein
A/Residues: 1-210, 'T', 212-394, 'D', 396-445, 'T', 447-700 <TMA>
A/Cross-references: GB:M23254; NID:9511636; PIDN:AA35645.1; PID:9511637
A/Note: parts of this sequence were determined by protein sequencing; the amino end o
J. Biol. Chem. 264, 6404-6411, 1989
R/Hata, A.; Ohno, S.; Akita, Y.; Suzuki, K.
A/Title: Tandemly reiterated negative enhancer-like elements regulate transcription o
A/Reference number: A33529; MIM:89197947; PMID:2539381
A/Accession: A33529
A/Molecule type: DNA
A/Residues: 1-67, 'G', 69-72, 'TE', 75-78, 'R' <HAT>
A/Cross-references: DDBJ:J04700; NID:9179910; PIDN:AA52760.1; PID:9463086
C/Genetics:
A/Gene: GDB:CAPN2; mCAPN; CAPNml
A/Cross-references: GDB:119750; OMIM:114230
A/Map position: 1pter-1qter
C/Complex: heterodimer of L (large) and S (small) chains
C/Function:
A/Description: catalyzes the hydrolysis of peptides
A/Note: cleaves preferentially after tyrosine, methionine, or arginine residues and b
C/Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic dom
C/Keywords: acetylated amino end; calcium binding; cysteine proteinase; EF hand; hete
F:2-700/Product: calpain large chain 2 #status predicted <MNT>
F:75-337/Domain: calpain catalytic domain homology <CALP>
F:529-560/Domain: calmodulin repeat homology <EF1>
F:572-604/Domain: calmodulin repeat homology <EF2>
F:605-634/Domain: calmodulin repeat homology <EF3>
F:637-669/Domain: calmodulin repeat homology <EF4>
F:2/Modified site: acetylated amino end (Ala) (in mature form) #status predicted
F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match 99.5%; Score 1287; DB 1; Length 700;
Best Local Similarity 99.6%; Pred. No. 5.2e-87;

Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINIREVILNRKFLPGEYILVPSFEEPNKDGDFCIRVFSEKKADY 60
 |||||
 Db 450 KNFFLTNRARERSDFINIREVILNRKFLPGEYILVPSFEEPNKDGDFCIRVFSEKKADY 509

QY 61 QAVDEIEANLEEFIDSEDDIDGFRRLFAQLAGEPAEISAFELQTLIRVLAKRODIDS 120
 |||||
 Db 510 QAVDEIEANLEEFIDSEDDIDGFRRLFAQLAGEPAEISAFELQTLIRVLAKRODIDS 569

QY 121 DGFSEITCKIMVMDLSDSGSKGLKEFYILMTKIOKYRIEIDVDRSGTMSYEMRK 180
 |||||
 Db 570 DGFSEITCKIMVMDLSDSGSKGLKEFYILMTKIOKYRIEIDVDRSGTMSYEMRK 629

QY 181 ALEAGFKMPCOLHQYIVARFADDDLIIDFNFVRLVLETLFKIFKOLDPEPNTGTIEL 240
 |||||
 Db 630 ALEAGFKMPCOLHQYIVARFADDDLIIDFNFVRLVLETLFKIFKOLDPEPNTGTIEL 689

QY 241 DLISWLCFSVL 251
 |||||
 Db 690 DLISWLCFSVL 700

RESULT 2

B24815
 calpain (EC 3.4.22.17) large chain 2 - rabbit (fragment)
 N:Alternate names: calcium-activated neutral proteinase (CANP)
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C>Date: 24-Jan-1988 #sequence_revision 24-Jan-1988 #text_change 16-Jul-1999
 C:Accession: B24815
 R:Morit, Y.; Kawasaki, H.; Sugihara, H.; Imajoh, S.; Kawashima, S.; Suzuki, K.
 J. Biol. Chem. 261, 9465-9471, 1986
 A:Title: Isolation and sequence analyses of cDNA clones for the large subunits of two is
 A:Reference number: A92594; MUID:86350902; PMID:2424911
 A:Accession: B24815
 A:Molecule type: mRNA
 A:Residues: 1-422 <EMO>
 A:Cross-references: GB:MI3797; NID:g165665; PIDN:AAA31455.1; PID:g165666
 C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
 C:Keywords: calcium binding; cysteine proteinase; EF hand; hydrolase
 F:251-282/Domain: calmodulin repeat homology <EF1>
 F:294-326/Domain: calmodulin repeat homology <EF2>
 F:327-356/Domain: calmodulin repeat homology <EF3>
 F:359-391/Domain: calmodulin repeat homology <EF4>

Query Match

95.2%; Score 1232; DB 2; Length 422;

Best Local Similarity 93.6%; Pred. No. 3.2e-83;

Matches 235; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINIREVILNRKFLPGEYILVPSFEEPNKDGDFCIRVFSEKKADY 60
 |||||
 Db 172 KNFFLTNRARERSDFINIREVILNRKFLPGEYILVPSFEEPNKDGDFCIRVFSEKKADY 231

QY 61 QAVDEIEANLEEFIDSEDDIDGFRRLFAQLAGEPAEISAFELQTLIRVLAKRODIDS 120
 |||||
 Db 232 QAVDEIEANLEEFIDSEDDIDGFRRLFAQLAGEPAEISAFELQTLIRVLAKRODIDS 291

QY 121 DGFSEITCKIMVMDLSDSGSKGLKEFYILMTKIOKYRIEIDVDRSGTMSYEMRK 180
 |||||
 Db 292 DGFSEITCKIMVMDLSDSGSKGLKEFYILMTKIOKYRIEIDVDRSGTMSYEMRK 351

QY 181 ALEAGFKMPCOLHQYIVARFADDDLIIDFNFVRLVLETLFKIFKOLDPEPNTGTIEL 240
 |||||
 Db 352 ALEAGFKMPCOLHQYIVARFADDDLIIDFNFVRLVLETLFKIFKOLDPEPNTGTIEL 411

QY 241 DLISWLCFSVL 251
 |||||
 Db 412 DLISWLCFSVL 422

RESULT 3

S38361
 calpain (EC 3.4.22.17) II large chain - rat

N:Alternate names: calpain II 80k chain
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 22-Jun-1999
 C:Accession: S38361; S08650; S39751
 R:Deluca, C.I.; Davies, P.L.; Samis, J.A.; Elce, J.S.
 Biochim. Biophys. Acta 1216, 81-93, 1993
 A:Title: Molecular cloning and bacterial expression of cDNA for rat calpain II 80 kDa
 A:Reference number: S38361; MUID:94032492; PMID:8218419
 A:Accession: S38361
 A:Molecule type: mRNA
 A:Residues: 1-700
 A:Cross-references: EMBL:L09120; NID:g402665; PIDN:AAA16327.1; PID:g402666
 R:Samis, J.A.; Back, D.W.; Graham, E.J.; Elce, J.S.
 submitted to the EMBL Data Library, February 1990
 A:Reference number: S08650
 A:Accession: S08650
 A:Molecule type: DNA
 A:Residues: 380-439 <SAM>
 A:Cross-references: EMBL:X51772
 C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic dom
 C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer;
 F:75-327/Domain: calpain catalytic domain homology <CALP>
 F:529-560/Domain: calmodulin repeat homology <EF1>
 F:572-604/Domain: calmodulin repeat homology <EF2>
 F:605-634/Domain: calmodulin repeat homology <EF3>
 F:637-669/Domain: calmodulin repeat homology <EF4>
 F:105,262,286/Active site: Cys, His, Asn #status predicted

Query Match

93.5%; Score 1210; DB 2; Length 700;

Best Local Similarity 92.4%; Pred. No. 2.4e-81;

Matches 232; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINIREVILNRKFLPGEYILVPSFEEPNKDGDFCIRVFSEKKADY 60
 |||||
 Db 450 KNFFLTNRARERSDFINIREVILNRKFLPGEYILVPSFEEPNKDGDFCIRVFSEKKADY 509

QY 61 QAVDEIEANLEEFIDSEDDIDGFRRLFAQLAGEPAEISAFELQTLIRVLAKRODIDS 120
 |||||
 Db 510 QAVDEIEANLEEFIDSEDDIDGFRRLFAQLAGEPAEISAFELQTLIRVLAKRODIDS 569

QY 121 DGFSEITCKIMVMDLSDSGSKGLKEFYILMTKIOKYRIEIDVDRSGTMSYEMRK 180
 |||||
 Db 570 DGFSEITCKIMVMDLSDSGSKGLKEFYILMTKIOKYRIEIDVDRSGTMSYEMRK 629

QY 181 ALEAGFKMPCOLHQYIVARFADDDLIIDFNFVRLVLETLFKIFKOLDPEPNTGTIEL 240
 |||||
 Db 630 ALEAGFKMPCOLHQYIVARFADDDLIIDFNFVRLVLETLFKIFKOLDPEPNTGTIEL 689

QY 241 DLISWLCFSVL 251
 |||||
 Db 690 DLISWLCFSVL 700

RESULT 4

S57194
 calpain (EC 3.4.22.17) large chain 2 - chicken
 N:Alternate names: m-calpain heavy chain
 C:Species: Gallus gallus (chicken)
 C>Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 C:Accession: S57194
 R:Sorimachi, H.; Tsukahara, T.; Okada-Ban, M.; Sugita, H.; Ishiura, S.; Suzuki, K.
 Biochim. Biophys. Acta 1261, 381-393, 1995
 A:Title: Identification of a third ubiquitous calpain species - chicken muscle expres
 A:Reference number: S57194; MUID:95260862; PMID:7742367
 A:Accession: S57194
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-700 <SOR>
 A:Cross-references: EMBL:D38026; NID:g882068; PIDN:BA07228.1; PID:g882069
 C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic dom
 C:Keywords: cysteine proteinase; EF hand; hydrolase
 F:75-327/Domain: calpain catalytic domain homology <CALP>
 F:529-560/Domain: calmodulin repeat homology <EF1>

OY 192 QLHOVAVAFADDDLIIDFDFVRCVRLVETLTKIFKQDLPENGTIELDLISWL 246
 DB 761 QLYDIITMRVADKHMNIDDFSCFVRLEGFRFAHFADKDGDIITKLVLEWL 815

RESULT 12

B34468
 calpain (EC 3.4.22.17) large chain 3 - rat
 N:Alternate names: cysteine proteinase
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: B34468; S10589
 R:Scrimachi, H.; Imaizumi-Ohmori, S.; Emori, Y.; Kawasaki, H.; Ohno, S.; Minami, Y.; Suzuki, J. Biol. Chem. 264, 20106-20111, 1989
 A:Title: Molecular cloning of a novel mammalian calcium-dependent protease distinct from A:Reference number: A34688; MUID:90062125; PMID:2555341
 A:Accession: B34468
 A:Molecule type: mRNA
 A:Residues: 1-821 <SOR>
 A:Cross-references: GB:J05121; NID:g205955; PIDN:AAA4190.1; PID:g205956
 R:Scrimachi, H.; Ohmi, S.; Emori, Y.; Kawasaki, H.; Saigo, T.C.; Ohno, S.; Minami, Y.; S Biol. Chem. Hoppe-Sevler 371(Suppl.), 171-176, 1990
 A:Title: A novel member of the calcium-dependent cysteine protease family.
 A:Reference number: S10589; MUID:90380278; PMID:2400579
 A:Accession: S10589
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-50, 'E', 52-211, 'V', 213-252, 'K', 254-821 <SO2>
 C:Superfamily: calpain large chain; calmodulin repeat homology; calpain catalytic domain
 C:Keywords: calcium binding; cysteine proteinase; EF hand; hydrolyase
 F:99-400/Domain: calpain catalytic domain homology <EF1>
 F:649-680/Domain: calmodulin repeat homology <EF2>
 F:692-724/Domain: calmodulin repeat homology <EF3>
 F:725-754/Domain: calmodulin repeat homology <EF4>
 F:757-789/Domain: calmodulin repeat homology <EF5>
 F:790-821/Domain: calmodulin repeat homology <EF5>
 F:129,334,358/Active site: Cys, His, Asn #status predicted

Query Match 47.1%; Score 609; DB 1; Length 821;
 Best Local Similarity 39.8%; Pred. No. 4.3e-37;
 Matches 117; Conservative 60; Mismatches 69; Indels 48; Gaps 3;

OY 1 KNFFLTNARERSDFILNREVLNFKLPGEYILVSTFEPNKKDGCIRYFSKK--- 57
 DB 522 KDFFLYNASKASKYTYIMNREVSOFRFLPSEYIVPSTYEPHOGGEFLRVSSEKRLNS 581
 OY 58 -----ADYQ-----AVDEIEANLE----- 73
 DB 582 EEAENTISVDRPVKKKKKPIIFVSDRANSKELGVDAEAGKRTGPKOGESPQRP 641
 OY 74 -FDISEDDIDGFRRLFAQLAGEDAEISAFELQTLRLVLAKKRODKSDGFSIETCKIIV 132
 DB 642 GHTDSESEEQOQFNIFRQIAGDMEICADELKNLVNVNKHKLDTQGFLESCRSMT 701
 OY 133 DMLDSGSGKGLKREYILMTKIKYKQKIRKIDVDRSGTMSYMKRALEAGKMKCQ 192
 DB 702 ALMDTDGGGRNLQEFHLMKKIKAMOKIFKHYPDHSCTINSYMRNAVNDAGHLNSQ 761
 OY 193 LHOVAVAFADDDLIIDFDFVRCVRLVETLTKIFKQDLPENGTIELDLISWL 246
 DB 762 LXDITTRVADKHMNIDDFSCFVRLEGFRFAHFADKDGDIITKLVLEWL 815

RESULT 13

A34466
 calpain (EC 3.4.22.17) light chain - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 16-Jul-1999
 C:Accession: A34466
 R:McClelland, P.; Lash, J.A.; Hathaway, D.R. J. Biol. Chem. 264, 17428-17431, 1989
 A:Title: Identification of major autolytic cleavage sites in the regulatory subunit of v

A:Reference number: A34466; MUID:9008911; PMID:2551902

A:Accession: A34466
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-263 <MCC>
 A:Cross-references: GB:J05065; NID:g162780; PIDN:AAA30422.1; PID:g162781
 C:Superfamily: calpain small chain; calmodulin repeat homology
 C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer;
 F:1-51/Domain: glycine-rich <GLY>
 F:91-122/Domain: calmodulin repeat homology <EF1>
 F:134-166/Domain: calmodulin repeat homology <EF2>
 F:167-199/Domain: calmodulin repeat homology <EF3>
 F:199-231/Domain: calmodulin repeat homology <EF4>
 F:232-263/Domain: calmodulin repeat homology <EF5>

Query Match 35.9%; Score 465; DB 2; Length 263;
 Best Local Similarity 44.7%; Pred. No. 4.3e-27;
 Matches 88; Conservative 42; Mismatches 67; Indels 0; Gaps 0;

OY 50 IRVSEKKADYQAVDEIEANLEEFDISDDIDGFRRLFAQLAGEDAEISAFELQTLR 109
 DB 61 ISATSEAAQYNPEVPPTHTYSNLEANESEVEVQFRFLFAQLAGDDMEVSATELMNLN 120
 OY 110 RYLAKRQDKSDGSIETCKIIVDMDSGSGKGLKREYILMTKIKYKQKIRKIDVDR 169
 DB 121 KVTTRHPDLKTDGIDICRSMVAVMSDITGKLGFEFFKLMNNIKWQAVYKQFVDR 180
 OY 170 SGTMSYMKRALEPAGKMPQCOLHOVAVAFADDDLIIDFDFVRCVRLVETLTKIFKQ 229
 DB 181 SGTGSSSLPCGAFPAFGRLEHLYNMILTRISDGGNMDPISCLVLDAMFPAFKS 240
 OY 230 LDPENTGTIELDLISWL 246
 DB 241 LDKQGTQIQVNIQEWL 257

RESULT 14

CIRGL

calpain (EC 3.4.22.17) small chain - pig
 N:Alternate names: calcium-activated neutral proteinase (CANP); calpain light chain;
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999

C:Accession: A25166; B25166
 R:Sakihama, T.; Kakidani, H.; Zenita, K.; Yumoto, N.; Kikuchi, T.; Sasaki, T.; Kannag Proc. Natl. Acad. Sci. U.S.A. 82, 6075-6079, 1985
 A:Title: A putative Ca²⁺-binding protein: structure of the light subunit of porcine c
 A:Reference number: A25166; MUID:85298299; PMID:2994060
 A:Accession: A25166
 A:Molecule type: mRNA
 A:Residues: 1-266 <SAK>
 A:Cross-references: GB:M11778; NID:g164402; PIDN:AAA31010.1; PID:g164403; GB:M11779;
 A:Accession: B25166
 A:Molecule type: protein
 A:Residues: 2-56;125-143;157-177;247-248;250-256;265-266 <SA2>
 C:Complex: heterodimer of L (large) and S (small) chains
 C:Function:
 A:Description: catalyzes the hydrolysis of peptides
 A>Note: cleaves preferentially after tyrosine, methionine, or arginine residues and b
 C:Superfamily: calpain small chain; calmodulin repeat homology
 C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; duplication;
 F:1-54/Domain: glycine-rich <GLY>
 F:94-125/Domain: calmodulin repeat homology <EF1>
 F:137-169/Domain: calmodulin repeat homology <EF2>
 F:170-199/Domain: calmodulin repeat homology <EF3>
 F:202-234/Domain: calmodulin repeat homology <EF4>
 F:235-266/Domain: calmodulin repeat homology <EF5>
 F:1/Modified site: acetylated amino end (Met) #status experimental

Query Match 35.9%; Score 465; DB 1; Length 266;
 Best Local Similarity 45.2%; Pred. No. 4.3e-27;
 Matches 89; Conservative 40; Mismatches 68; Indels 0; Gaps 0;

OY 50 IRVSEKKADYQAVDEIEANLEEFDISDDIDGFRRLFAQLAGEDAEISAFELQTLR 109

```
Db      64  ISAISEAAQYNPEPPPRTHYSNIANESEVROFRRLFAQLAGDDMEVSATLMTLN 123
OY      110  RVLAKRODIKSDGFSIECKIMVMDLSDSGKLGKFEFFILMTKIQKQKTYREIDYR 169
Db      124  KVVTRHPDLKTDGIDTCRSMVAVMDSDTGKLGFEFFKYLWNNIKKWAIFYKQFDYDR 183
OY      170  SGTNNSYEMRKALEEAGFKMPCOLHOVIVARFADQLIIDPFVRCIVRETLFKIRKQ 229
Db      184  SGTIGSSSLPGAFEAAGFLNHLNMLIRYSDSGNMDFNFIISCLVRLDAMFRARKS 243
OY      230  LDPENTGTIELDLISWL 246
Db      244  LDKDGTGQIQVNIQEWL 260
```

RESULT 15

```
CTRBL
calpain (EC 3.4.22.17) small chain - rabbit
N:Alternate names: calcium-activated neutral proteinase (CANP); calpain light chain; cal
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
C:Accession: A24816
R:Emori, Y.; Kawasaki, H.; Imajob, S.; Kawashima, S.; Suzuki, K.
J. Biol. Chem. 261, 9472-9476, 1986
A:Title: Isolation and sequence analysis of cDNA clones for the small subunit of rabbit
A:Reference number: A24816; MUID:86250903; PMID:3013892
A:Accession: A24816
A:Molecule type: mRNA
A:Residues: 1-266 <EMO>
A:Cross-references: GB:M1364; NID:9164875; PIDN:AAA81565.1; PID:9164876
C:Complex: heterodimer of L (large) and S (small) chains
C:Function:
A:Description: catalyzes the hydrolysis of peptides
A>Note: Cleaves preferentially after tyrosine, methionine, or arginine residues and bef
C:Superfamily: calpain small chain; calmodulin repeat homology
C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer; hyc
F:1-54/Domain: glycine-rich <GLY>
F:94-125/Domain: calmodulin repeat homology <EF1>
F:137-169/Domain: calmodulin repeat homology <EF2>
F:170-199/Domain: calmodulin repeat homology <EF3>
F:202-234/Domain: calmodulin repeat homology <EF4>
F:235-266/Domain: calmodulin repeat homology <EF5>
```

```
Query Match      35.6%; Score 461; DB 1; Length 266;
Best Local Similarity 44.7%; Pred. No. 8.5e-27;
Matches 88; Conservative 41; Mismatches 68; Indels 0; Gaps 0;
```

```
OY      50  IRVSEKKADYQAVDELEIANLEEFDISDDIDGFRRLFAQLAGEDAISAFELQTLIR 109
Db      64  ISAISEAAQYNPEPPPRTHYSNIANESEVROFRRLFAQLAGDDMEVSATLMTLN 123
OY      110  RVLAKRODIKSDGFSIECKIMVMDLSDSGKLGKFEFFILMTKIQKQKTYREIDYR 169
Db      124  KVVTRHPDLKTDGIDTCRSMVAVMDSDTGKLGFEFFKYLWNNIKKWAIFYKQFDYDR 183
OY      170  SGTNNSYEMRKALEEAGFKMPCOLHOVIVARFADQLIIDPFVRCIVRETLFKIRKQ 229
Db      184  SGTIGSSSLPGAFEAAGFLNHLNMLIRYSDSGNMDFNFIISCLVRLDAMFRARKS 243
OY      230  LDPENTGTIELDLISWL 246
Db      244  LDKDGTGQIQVNIQEWL 260
```

Search completed: July 24, 2003, 12:53:46
Job time : 14.3158 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:53:55 ; Search time 16.9993 Seconds
(without alignments)
1753.524 Million cell updates/sec

Title: US-09-884-319a-6
Perfect score: 1294
Sequence: 1 KNFFLTNRARSDFTFNLRLPGEYILVPSFTPEPKDDFCIRVSEKKADY 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1294	100.0	251	9	US-09-840-707A-8
2	1294	100.0	251	11	US-09-884-319-6
3	1294	100.0	251	11	US-10-038-557A-8
4	1287	99.5	700	9	US-09-840-707A-9
5	1287	99.5	700	10	US-09-768-877-23
6	1287	99.5	700	11	US-09-884-319-7
7	1287	99.5	700	15	US-10-038-557A-9
8	1287	99.5	700	15	US-10-116-519-12
9	1051	81.2	206	15	US-10-043-487-221
10	773.5	58.5	713	15	US-10-116-519-8
11	757.5	58.5	714	10	US-09-768-877-22
12	757.5	58.5	714	15	US-10-116-519-6
13	646.5	50.0	702	15	US-10-116-519-11
14	633.5	49.0	703	10	US-09-768-877-27
15	624	48.2	709	15	US-10-116-519-9

16	613	47.4	821	10	US-09-768-877-24	Sequence 24, Appl
17	613	47.4 <td>821</td> <td>15<td>US-10-116-519-7</td><td>Sequence 7, Appl</td></td>	821	15 <td>US-10-116-519-7</td> <td>Sequence 7, Appl</td>	US-10-116-519-7	Sequence 7, Appl
18	597.5	46.2	664	15 <td>US-10-325-881-70</td> <td>Sequence 70, Appl</td>	US-10-325-881-70	Sequence 70, Appl
19	597.5	46.2	690	10	US-09-768-877-18	Sequence 28, Appl
20	597.5	46.2	690	15 <td>US-10-116-519-13</td> <td>Sequence 13, Appl</td>	US-10-116-519-13	Sequence 13, Appl
21	470	36.3	123	9	US-09-925-297-527	Sequence 527, Appl
22	465	35.9	207	9	US-09-925-301-1001	Sequence 1001, Ap
23	240	18.5	198	10	US-09-981-353-138	Sequence 138, Ap
24	234	18.1	217	11	US-09-919-039-64	Sequence 64, Appl
25	232	17.9	92	15	US-10-106-698-5122	Sequence 5122, Ap
26	225.5	17.4	694	15	US-10-116-519-54	Sequence 54, Appl
27	225.5	17.4	697	15	US-10-116-519-56	Sequence 56, Appl
28	191	14.8	669	10	US-09-962-739-2	Sequence 490, Appl
29	188	14.5	74	15	US-10-177-293-490	Sequence 62, Appl
30	185.5	14.3	284	14	US-10-052-586-236	Sequence 236, Appl
31	185.5	14.3	284	14	US-10-063-547-62	Sequence 62, Appl
32	185.5	14.3	284	14	US-10-174-590-236	Sequence 236, Appl
33	185.5	14.3	284	15	US-10-176-758-236	Sequence 236, Appl
34	185.5	14.3	284	15	US-10-175-737-236	Sequence 236, Appl
35	185.5	14.3	284	15	US-10-063-616-62	Sequence 62, Appl
36	185.5	14.3	284	15	US-10-173-706-236	Sequence 236, Appl
37	185.5	14.3	284	15	US-10-173-736-236	Sequence 236, Appl
38	185.5	14.3	284	15	US-10-175-752-236	Sequence 236, Appl
39	185.5	14.3	284	15	US-10-176-482-236	Sequence 236, Appl
40	185.5	14.3	284	15	US-10-176-757-236	Sequence 236, Appl
41	185.5	14.3	284	15	US-10-176-913-236	Sequence 236, Appl
42	185.5	14.3	284	15	US-10-180-552-236	Sequence 236, Appl
43	185.5	14.3	284	15	US-10-180-557-236	Sequence 236, Appl
44	185.5	14.3	284	15	US-10-063-502-62	Sequence 62, Appl
45	185.5	14.3	284	15	US-10-063-502-62	Sequence 62, Appl

ALIGNMENTS

RESULT 1
US-09-840-707A-8
; Sequence 8, Application US/09840707A
; Patent No. US20020077276A1
; GENERAL INFORMATION:
; APPLICANT: Friedeking, Terry M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
; FILE REFERENCE: 24881-301C
; CURRENT APPLICATION NUMBER: US/09/840, 707A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/562, 979
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/198, 210
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: IL-1 receptor intracellular ligand protein
; OTHER INFORMATION: comprising amino acid sequence
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5, 817, 476
; PATENT FILING DATE: 1995-06-07
; PUBLICATION DATE: 1998-10-06
US-09-840-707A-8

Query Match 100.0%; Score 1294; DB 9; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.7e-124;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KNFFLTNRARSDFTFNLRLPGEYILVPSFTPEPKDDFCIRVSEKKADY 60
DB 1 KNFFLTNRARSDFTFNLRLPGEYILVPSFTPEPKDDFCIRVSEKKADY 60

QY	QAVDDEIANNEEFIISEDDIDDGRRRLPAOLAGEAEISAFLOTILRRVLAKRODIX 120
QY	61 QAVDDEIANNEEFIISEDDIDDGRRRLPAOLAGEAEISAFLOTILRRVLAKRODIX 120
Db	61 QAVDDEIANNEEFIISEDDIDDGRRRLPAOLAGEAEISAFLOTILRRVLAKRODIX 120
QY	121 DGFSEIETCKIWMMDLSDSGSKGLKEFYILWTKLOKYOKIYREIDVDSGTMNSYEMRK 180
QY	121 DGFSEIETCKIWMMDLSDSGSKGLKEFYILWTKLOKYOKIYREIDVDSGTMNSYEMRK 180
Db	121 DGFSEIETCKIWMMDLSDSGSKGLKEFYILWTKLOKYOKIYREIDVDSGTMNSYEMRK 180
QY	181 ALEEGGFMMPCOLHOVIYARFADDOILIDFQNFVRCVLVLETLFKIFKOLDPENTGTIEL 240
QY	181 ALEEGGFMMPCOLHOVIYARFADDOILIDFQNFVRCVLVLETLFKIFKOLDPENTGTIEL 240
Db	181 ALEEGGFMMPCOLHOVIYARFADDOILIDFQNFVRCVLVLETLFKIFKOLDPENTGTIEL 240
QY	241 DLISWLCFSYVL 251
QY	241 DLISWLCFSYVL 251
Db	241 DLISWLCFSYVL 251

```

RESULT 2
US-09-884-319-6
Sequence 6, Application US/09884319
Publication No. US20030124625A1
GENERAL INFORMATION:
APPLICANT: Lin, Lin-Ling
Graham, James
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
BINDING
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/884,319
FILING DATE: 18-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/083,516
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/487,942
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8324
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 251 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-884-319-6
Query Match 100.0%; Score 1294; DB II; Length 251;
Best Local Similarity 100.0%; Pred. No. 1,7e-124;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 KNFPLTNAREKSDPFINLRVLRKRLPGGEYILVPSSTPEPNKDGDFCIKRVSEKKADY 60
|||||
1 KNFPLTNAREKSDPFINLRVLRKRLPGGEYILVPSSTPEPNKDGDFCIKRVSEKKADY 60

```

QY	6	QAADDELEANLEEFDISEDDIDGRRRLFAOLAGDAEISAFELQTLIRLAKRODKS	120
Db	61	QAADDELEANLEEFDISEDDIDGRRRLFAOLAGDAEISAFELQTLIRLAKRODKS	120
QY	121	DGFSEICTKIMVMDMLSDSGSLGLKEFYILMTKIQOKIYREIDVRSGTMMYSMRK	180
Db	121	DGFSEICTKIMVMDMLSDSGSGKLGLKEFYILMTKIQOKIYREIDVRSGTMMYSMRK	180
QY	181	AEEAEFGKMPCCOLHOVIYARFADDOLIIDPDMFVACLVRETLTKIKOLDPENTGTIEL	240
Db	181	AEEAEFGKMPCCOLHOVIYARFADDOLIIDPDMFVACLVRETLTKIKOLDPENTGTIEL	240
QY	241	DLISWLCFSYL	251
Db	241	DLISWLCFSYL	251

```

RESULT 3
US-10-038-557A-8
; Sequence 8, Application US/10038557A
; Publication No. US20030092684A1
; GENERAL INFORMATION:
; APPLICANT: Ignatyev, George M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
; TITLE OF INVENTION: INFECTIONS AND OTHER DISORDERS
; FILE REFERENCE: 24681-301D
; CURRENT APPLICATION NUMBER: US/10/038, 557A
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/840,707
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/562,979
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/198,210
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: IL-1 receptor Intracellular ligand protein
; OTHER INFORMATION: comprising amino acid sequence
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5,817,476
; PATENT FILING DATE: 1995-06-07
; PUBLICATION DATE: 1998-10-06
US-10-038-557A-8

Query Match      100.0%; Score 1294; DB 15; Length 251;
Best Local Similarity 100.0%; Pred. No. 1,7e124;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1  KNFELTNARRSRDPFIMLRVNLNFKLPPEGVILVPTSEPNKGGDCPIRVSEKKADY 60
      |||
Db      1  KNFELTNARRSRDPFIMLRVNLNFKLPPEGVILVPTSEPNKGGDCPIRVSEKKADY 60
QY      61  QAVDDIEANLEEFDISDDIDDGFRRLPAQLAGEDAETISAFELQTLIRVLAKRQDKS 120
      |||
Db      61  QAVDDIEANLEEFDISDDIDDGFRRLPAQLAGEDAETISAFELQTLIRVLAKRQDKS 120
QY      121  DGSFETNCKIIVMDMDSGSGKLGKEFYIIMTKIQKQOKIYREIDVDSGTMNSYEMRK 180
      |||
Db      121  DGSFETNCKIIVMDMDSGSGKLGKEFYIIMTKIQKQOKIYREIDVDSGTMNSYEMRK 180
QY      181  ALEEGEFKMPQLHGVIVARFADQDLIDFNFVACLVRLETLFKIFKOLDPENTGTIEL 240
      |||
Db      181  ALEEGEFKMPQLHGVIVARFADQDLIDFNFVACLVRLETLFKIFKOLDPENTGTIEL 240
QY      241  DLISMLCFSVL 251
      |||

```

Db 241 DLISWLCFSVL 251

RESULT 4

US-09-840-707A-9

; Sequence 9, Application US/09840707A

; Patent No. US2002007276A1

; GENERAL INFORMATION:

; APPLICANT: Fredeking, Terry M.

; APPLICANT: Ignatyev, George M.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS

; FILE REFERENCE: 24881-301C

; CURRENT APPLICATION NUMBER: US/09/840,707A

; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: 09/562,979

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 60/198,210

; PRIOR FILING DATE: 1999-04-27

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 700

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: IL-1 receptor intracellular ligand protein

; OTHER INFORMATION: comprising amino acid sequence

; PUBLICATION INFORMATION:

; PATENT DOCUMENT NUMBER: 5,817,476

; PATENT FILING DATE: 1995-06-07

; PUBLICATION DATE: 1998-10-06

US-09-840-707A-9

Query Match 99.5%; Score 1287; DB 9; Length 700;

Best Local Similarity 99.6%; Pred. No. 3.7e-123;

Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINLREVLNRFKLPGEYILVPSTFEPNKGDFCIRFSEKKADY 60

DB 450 KNFFLTNRARERSDFINLREVLNRFKLPGEYILVPSTFEPNKGDFCIRFSEKKADY 509

QY 61 QAVDEIEANLEEFDISDDIDGFRRLFAQLAGDAEISAFELQTIIRVLAKEODIKS 120

DB 510 QAVDEIEANLEEFDISDDIDGFRRLFAQLAGDAEISAFELQTIIRVLAKEODIKS 569

QY 121 DGFSTETCKINWMDSDSGSKGLGKEYIILMTKIQOKYIREIDVDVRSGMTNSYEMRK 180

DB 570 DGFSTETCKINWMDSDSGSKGLGKEYIILMTKIQOKYIREIDVDVRSGMTNSYEMRK 629

QY 181 ALEAGFKMPCQLHOVIYARFADQLIIDFNFVACLVRLFTLFRIFQOLDPENTGTIEL 240

DB 630 ALEAGFKMPCQLHOVIYARFADQLIIDFNFVACLVRLFTLFRIFQOLDPENTGTIEL 689

QY 241 DLISWLCFSVL 251

DB 690 DLISWLCFSVL 700

RESULT 5

US-09-768-877-23

; Sequence 23, Application US/09768877

; Patent No. US20020150896A1

; GENERAL INFORMATION:

; APPLICANT: POLONSKY, KENNETH S.

; APPLICANT: HORIKAWA, YUKIO

; APPLICANT: ODA, NAOHISA

; APPLICANT: COX, NANCY J.

; APPLICANT: SEEMAN, SEAMUS

; APPLICANT: ZHOU, YUN-PING

; APPLICANT: OTANI, KENICHI

; APPLICANT: HANIS, CRAIG I.

; APPLICANT: BELL, GRAEME I.

; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES

; FILE REFERENCE: ARCD:307

; CURRENT APPLICATION NUMBER: US/09/768,877

; CURRENT FILING DATE: 2001-01-23

; PRIOR APPLICATION NUMBER: 09/422,869

; PRIOR FILING DATE: 1999-10-21

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 23

; LENGTH: 700

; TYPE: PRT

; ORGANISM: Human

US-09-768-877-23

Query Match 99.5%; Score 1287; DB 10; Length 700;

Best Local Similarity 99.6%; Pred. No. 3.7e-123;

Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINLREVLNRFKLPGEYILVPSTFEPNKGDFCIRFSEKKADY 60

DB 450 KNFFLTNRARERSDFINLREVLNRFKLPGEYILVPSTFEPNKGDFCIRFSEKKADY 509

QY 61 QAVDEIEANLEEFDISDDIDGFRRLFAQLAGDAEISAFELQTIIRVLAKEODIKS 120

DB 510 QAVDEIEANLEEFDISDDIDGFRRLFAQLAGDAEISAFELQTIIRVLAKEODIKS 569

QY 121 DGFSTETCKINWMDSDSGSKGLGKEYIILMTKIQOKYIREIDVDVRSGMTNSYEMRK 180

DB 570 DGFSTETCKINWMDSDSGSKGLGKEYIILMTKIQOKYIREIDVDVRSGMTNSYEMRK 629

QY 181 ALEAGFKMPCQLHOVIYARFADQLIIDFNFVACLVRLFTLFRIFQOLDPENTGTIEL 240

DB 630 ALEAGFKMPCQLHOVIYARFADQLIIDFNFVACLVRLFTLFRIFQOLDPENTGTIEL 689

QY 241 DLISWLCFSVL 251

DB 690 DLISWLCFSVL 700

RESULT 6

US-09-884-319-7

; Sequence 7, Application US/09884319

; Publication No. US20030124625A1

; GENERAL INFORMATION:

; APPLICANT: Lin, Lih-Ling

; APPLICANT: Graham, James

; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR

; BINDING INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; STREET: 87 Cambridgepark Drive

; CITY: Cambridge

; STATE: MA

; COUNTRY: USA

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/884,319

; FILING DATE: 18-Jun-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/083,516

; FILING DATE: <Unknown>

; APPLICATION NUMBER: 08/487,942

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-884-319-7

Query Match 99.5%; Score 1287; DB 11; Length 700;
Best Local Similarity 99.6%; Pred. No. 3.7e-123;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFELTNARERSDFINMLREVLNRFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 60
DB 450 KNFELTNARERSDFINMLREVLNRFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 509
QY 61 QAVDEIFANLEEDPISDDIDGFRRLPAQLAGDAEISAFELQTIIRVLAQRDIDS 120
DB 510 QAVDEIFANLEEDPISDDIDGFRRLPAQLAGDAEISAFELQTIIRVLAQRDIDS 569
QY 121 DGFSETECKIWMMDSDSGSKGLGKEYIILMTKIOKYOKIYREIDVDRSGTMSYEMRK 180
DB 570 DGFSETECKIWMMDSDSGSKGLGKEYIILMTKIOKYOKIYREIDVDRSGTMSYEMRK 629
QY 181 ALEAGFMPQOLHGVIVARFADQLIIDFNFVACLVRLLETKIFKQDPENGTIEL 240
DB 630 ALEAGFMPQOLHGVIVARFADQLIIDFNFVACLVRLLETKIFKQDPENGTIEL 689
QY 241 DLISWLCFSVL 251
DB 690 DLISWLCFSVL 700

RESULT 7
US-10-038-557A-9
Sequence 9, Application US/10038557A
Publication No. US20030092684A1
GENERAL INFORMATION:
APPLICANT: Ignat'yev, Terry M.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING HEMORRHAGIC VIRUS
FILE REFERENCE: 24881-301D
CURRENT APPLICATION NUMBER: US/10/038,557A
PRIOR FILING DATE: 2002-06-05
PRIOR APPLICATION NUMBER: 09/840,707
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: 09/562,979
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/198,210
PRIOR FILING DATE: 1999-04-27
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 700
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: IL-1 receptor intracellular ligand protein
PUBLICATION INFORMATION:
PATENT DOCUMENT NUMBER: 5,817,476
PATENT FILING DATE: 1995-06-07
PUBLICATION DATE: 1998-10-06
US-10-038-557A-9

Query Match 99.5%; Score 1287; DB 15; Length 700;
Best Local Similarity 99.6%; Pred. No. 3.7e-123;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFELTNARERSDFINMLREVLNRFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 60
DB 450 KNFELTNARERSDFINMLREVLNRFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 509
QY 61 QAVDEIFANLEEDPISDDIDGFRRLPAQLAGDAEISAFELQTIIRVLAQRDIDS 120
DB 510 QAVDEIFANLEEDPISDDIDGFRRLPAQLAGDAEISAFELQTIIRVLAQRDIDS 569
QY 121 DGFSETECKIWMMDSDSGSKGLGKEYIILMTKIOKYOKIYREIDVDRSGTMSYEMRK 180
DB 570 DGFSETECKIWMMDSDSGSKGLGKEYIILMTKIOKYOKIYREIDVDRSGTMSYEMRK 629
QY 181 ALEAGFMPQOLHGVIVARFADQLIIDFNFVACLVRLLETKIFKQDPENGTIEL 240
DB 630 ALEAGFMPQOLHGVIVARFADQLIIDFNFVACLVRLLETKIFKQDPENGTIEL 689
QY 241 DLISWLCFSVL 251
DB 690 DLISWLCFSVL 700

RESULT 8
US-10-116-519-12
Sequence 12, Application US/10116519
Publication No. US20030114373A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPA
FILE REFERENCE: D0124 NP
CURRENT APPLICATION NUMBER: US/10/116,519
PRIOR FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: US 60/281,253
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: US 60/288,768
PRIOR FILING DATE: 2001-05-04
PRIOR APPLICATION NUMBER: US 60/296,180
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/300,620
PRIOR FILING DATE: 2001-06-25
NUMBER OF SEQ ID NOS: 145
SOFTWARE: PatentIn version 3.0
SEQ ID NO 12
LENGTH: 700
TYPE: PRT
ORGANISM: Homo sapiens
US-10-116-519-12

Query Match 99.5%; Score 1287; DB 15; Length 700;
Best Local Similarity 99.6%; Pred. No. 3.7e-123;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFELTNARERSDFINMLREVLNRFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 60
DB 450 KNFELTNARERSDFINMLREVLNRFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 509
QY 61 QAVDEIFANLEEDPISDDIDGFRRLPAQLAGDAEISAFELQTIIRVLAQRDIDS 120
DB 510 QAVDEIFANLEEDPISDDIDGFRRLPAQLAGDAEISAFELQTIIRVLAQRDIDS 569
QY 121 DGFSETECKIWMMDSDSGSKGLGKEYIILMTKIOKYOKIYREIDVDRSGTMSYEMRK 180
DB 570 DGFSETECKIWMMDSDSGSKGLGKEYIILMTKIOKYOKIYREIDVDRSGTMSYEMRK 629
QY 181 ALEAGFMPQOLHGVIVARFADQLIIDFNFVACLVRLLETKIFKQDPENGTIEL 240
DB 630 ALEAGFMPQOLHGVIVARFADQLIIDFNFVACLVRLLETKIFKQDPENGTIEL 689

QY 241 DLISWLCFSVL 251
| | | | |
DB 690 DLISWLCFSVL 700

RESULT 9

US-10-043-487-221
; Sequence 221, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: PIERRE, LIGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella flexneri polypeptid
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 221
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Shigella flexneri
US-10-043-487-221

Query Match 81.2%; Score 1051; DB 15; Length 206;
Best Local Similarity 99.5%; Pred. No. 1e-99; Indels 0; Gaps 0;

Matches 203; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 46 GDFCIRVSEKKADYQAVDEIEANIEEFDDISEDDIDGFRLLFAQLAGEDAEISAFELQ 105
| | | | |
DB 1 GDFCIRVSEKKADYQAVDEIEANIEEFDDISEDDIDGFRLLFAQLAGEDAEISAFELQ 60
| | | | |
QY 106 TIRRVYLAQRDQKSDGSIETCKIMVMDSDSGSKGLKFEYILMTKIQYKYYREI 165
| | | | |
DB 61 TIRRVYLAQRDQKSDGSIETCKIMVMDSDSGSKGLKFEYILMTKIQYKYYREI 120
| | | | |
QY 166 DVDRSGTNSYEMRKALKEAGFKMPCQLHOVYVARFADQDLIDPNDNFVRCVRLTETLFK 225
| | | | |
DB 121 DVDRSGTNSYEMRKALKEAGFKMPCQLHOVYVARFADQDLIDPNDNFVRCVRLTETLFK 180
| | | | |
QY 226 IFKQDPENTGTIELDLISWLCFSVL 251
| | | | |
DB 181 IFKQDPENTGTIELDLISWLCFSVL 206
| | | | |

RESULT 10

US-10-116-519-8
; Sequence 8, Application US/10116519
; Publication No. US20030114373A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPAIN
; FILE REFERENCE: D0124 NP
; CURRENT APPLICATION NUMBER: US/10/116,519
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: US 60/281,253
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/288,768
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/296,180
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/300,620
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Mus musculus

US-10-116-519-8

Query Match 59.8%; Score 773.5; DB 15; Length 713;
Best Local Similarity 56.7%; Pred. No. 1.5e-70;
Matches 140; Conservative 58; Mismatches 48; Indels 1; Gaps 1;

QY 1 KNFFLTNRARERSDFINIREVLNRFKLPPEGYIIVPSTFEPNKGDFCIRVSEKKADY 60
| | | | |
DB 461 RQFLANASRAQSEHFINIREVSNRIKRLPPEGYIIVPSTFEPNKGDFLRFSEKSGT 520
| | | | |
QY 61 QAVDEIEANL-EEFDISEDDIDGFRLLFAQLAGEDAEISAFELQTLIRVYLAQRDQK 119
| | | | |
DB 521 QELDDQIQANLPDEKVLSEEDIDNFKTLFSKLAGDMEISVKEHQTILNRIISKHDLR 580
| | | | |
QY 120 SNGFSIETCKIIVMDLSDSGSKGLKEFYILMTKIQYKYYREIDVDRSGTNSYEMR 179
| | | | |
DB 581 TNGFSIESCRSMVNLMDRGNKGLTVEENILMNRIYLTIFRFEDLDKSGSMSAYEMR 640
| | | | |
QY 180 KALEAGFKMPCQLHOVYVARFADQDLIDPNDNFVRCVRLTETLFKIFKQDPENTGTIE 239
| | | | |
DB 641 MALEAGFKLNKKLHEILITRSEPDVAVDFNFCVCLVRLTETLMTFRFKLDTDLGGVY 700
| | | | |
QY 240 LDLSWL 246
| | | | |
DB 701 FDLFKWL 707
| | | | |

RESULT 11

US-09-768-877-22
; Sequence 22, Application US/09768877
; Patent No. US20020150896A1
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: GREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/768,877
; CURRENT FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 09/422,869
; PRIOR FILING DATE: 1995-10-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Human
US-09-768-877-22

Query Match 58.5%; Score 757.5; DB 10; Length 714;
Best Local Similarity 55.5%; Pred. No. 6.7e-69;
Matches 137; Conservative 57; Mismatches 52; Indels 1; Gaps 1;

QY 1 KNFFLTNRARERSDFINIREVLNRFKLPPEGYIIVPSTFEPNKGDFCIRVSEKKADY 60
| | | | |
DB 462 RQFLANASRAQSEHFINIREVSNRIKRLPPEGYIIVPSTFEPNKGDFLRFSEKSGT 521
| | | | |
QY 61 QAVDEIEANL-EEFDISEDDIDGFRLLFAQLAGEDAEISAFELQTLIRVYLAQRDQK 119
| | | | |
DB 522 VELDDQIQANLPDEKVLSEEDIDNFKTLFSKLAGDMEISVKEHQTILNRIISKHDLR 581
| | | | |
QY 120 SNGFSIETCKIIVMDLSDSGSKGLKEFYILMTKIQYKYYREIDVDRSGTNSYEMR 179
| | | | |
DB 582 TNGFSIESCRSMVNLMDRGNKGLTVEENILMNRIYLTIFRFEDLDKSGSMSAYEMR 641
| | | | |
QY 180 KALEAGFKMPCQLHOVYVARFADQDLIDPNDNFVRCVRLTETLFKIFKQDPENTGTIE 239
| | | | |

Db 642 MAISAGFKLNKKLYELITRYSEBPLAVDFDNFVCLVLETMERFFKTLDTDIDGVT 701
 QY 240 LDLISWL 246
 Db 702 FDLFKWL 708

RESULT 12

US-10-116-519-6
 ; Sequence 6, Application US/10116519
 ; Publication No. US20030114373A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company
 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPAIN
 FILE REFERENCE: D0124 NP
 CURRENT APPLICATION NUMBER: US/10/116,519

PRIOR FILING DATE: 2002-04-03
 PRIOR APPLICATION NUMBER: US 60/281,253
 PRIOR FILING DATE: 2001-04-03
 PRIOR APPLICATION NUMBER: US 60/288,768
 PRIOR FILING DATE: 2001-05-04
 PRIOR APPLICATION NUMBER: US 60/296,180
 PRIOR FILING DATE: 2001-06-06
 PRIOR APPLICATION NUMBER: US 60/300,620
 PRIOR FILING DATE: 2001-06-25
 NUMBER OF SEQ ID NOS: 145
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 6
 LENGTH: 714
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-116-519-6

Query Match 58.5%; Score 757.5; DB 15; Length 714;
 Best Local Similarity 55.5%; Pred. No. 6,7e-69;
 Matches 137; Conservative 57; Mismatches 52; Indels 1; Gaps 1;

QY 1 KNFELTNARERSDFTINLRVNLNFKLPGEYILVPSFTEPNKDGDCIRVSEKKADY 60
 Db 462 RDFELANASRARSSEQFINLRVSTFRRLPGEYVVPSTFEPNKGSDVLRFSSEKSGT 521
 QY 61 QAVDEITANL-EEFDISEDDIDGFRRLFAQLAGEDAISAFLQTLIRVLAKRODK 119
 Db 522 VELDQIQANLPDEQVLSSEEDIDENFKALFQLAGEDAEISVKELRTILNRIISKHDLR 581
 QY 120 SDGSIETCKTMDMLSDSGSGKLGKEFYILMTKIQYKQIYREIDVDSGTMSYEMR 179
 Db 582 TKGFSLESCRSVMNLMDDGNGKGLGVFNILMNIRIYLSIFRFDLKSGSMAYEMR 641
 QY 180 KALEBAGFKMPCQLHQVIVARFADQOLIIDFDFNFRCLVRLTETLEKIFKQOLDPENTGIE 239
 Db 642 MAIESAGFKLNKKLYELITRYSEBPLAVDFDNFVCLVLETMERFFKTLDTDIDGVT 701
 QY 240 LDLISWL 246
 Db 702 FDLFKWL 708

RESULT 13

US-10-116-519-11
 ; Sequence 11, Application US/10116519
 ; Publication No. US20030114373A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company
 TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPAIN
 FILE REFERENCE: D0124 NP
 CURRENT APPLICATION NUMBER: US/10/116,519

PRIOR FILING DATE: 2002-04-03
 PRIOR APPLICATION NUMBER: US 60/281,253
 PRIOR FILING DATE: 2001-04-03
 PRIOR APPLICATION NUMBER: US 60/288,768

;; PRIOR FILING DATE: 2001-05-04
 ; PRIOR APPLICATION NUMBER: US 60/296,180
 ; PRIOR FILING DATE: 2001-06-06
 ; PRIOR APPLICATION NUMBER: US 60/300,620
 ; PRIOR FILING DATE: 2001-06-25
 ; NUMBER OF SEQ ID NOS: 145
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 11
 ; LENGTH: 702
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-116-519-11

Query Match 50.0%; Score 646.5; DB 15; Length 702;
 Best Local Similarity 49.4%; Pred. No. 1.5e-57;
 Matches 122; Conservative 50; Mismatches 74; Indels 1; Gaps 1;

QY 1 KNFELTNARERSDFTINLRVNLNFKLPGEYILVPSFTEPNKDGDCIRVSEKKADY 60
 Db 450 KEFFTKYODHGFSEIFTSREVSSQLRPGEYIITPSTFEPNHRADLFLVFEKSES 509
 QY 61 QAVDD-EIEANLEEFDISDDIDGFRRLFAQLAGEDAISAFLQTLIRVLAKRODK 119
 Db 510 WELDEVNNAEQLQEKVSEDDMDQFLHFKIVAGEGKIGVYELQRLNLMATKKSFK 569
 QY 120 SDGSIETCKTMDMLSDSGSGKLGKEFYILMTKIQYKQIYREIDVDSGTMSYEMR 179
 Db 570 TKGFGLDACRCMINLMDDGSGKLGLEFKYILMTKIKKWMIDIFRECDDHSGTINSYEMR 629
 QY 180 KALEBAGFKMPCQLHQVIVARFADQOLIIDFDFNFRCLVRLTETLEKIFKQOLDPENTGIE 239
 Db 630 LVIEKAGIKLNKKVQVIVARFADQOLIIDFDFSFISCFRLKMTTFELVMDPKVTHIC 689
 QY 240 LDLISWL 246
 Db 690 LSLQWL 696

RESULT 14

US-09-768-877-27
 ; Sequence 27, Application US/09768877
 ; Patent No. US20020150896A1

GENERAL INFORMATION:

APPLICANT: POLONSKY, KENNETH S.
 APPLICANT: HORIKAWA, YUKIO
 APPLICANT: ODA, NAOHISA
 APPLICANT: COX, NANCY J.
 APPLICANT: SREENAN, SEAMUS
 APPLICANT: ZHOU, YUN-PING
 APPLICANT: OTANI, KENICHI
 APPLICANT: HANIS, CRAIG L.
 APPLICANT: BELT, GRAEME I.
 TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
 FILE REFERENCE: ARCD:307
 CURRENT APPLICATION NUMBER: US/09/768,877-
 PRIOR FILING DATE: 2001-01-23
 PRIOR APPLICATION NUMBER: 09/422,869
 PRIOR FILING DATE: 1999-10-21
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 27
 LENGTH: 703
 TYPE: PRT
 ORGANISM: RAT
 US-09-768-877-27

US-09-768-877-27

Query Match 49.0%; Score 633.5; DB 10; Length 703;
 Best Local Similarity 48.4%; Pred. No. 3.3e-56;
 Matches 121; Conservative 54; Mismatches 70; Indels 5; Gaps 2;

QY 1 KNFELTNARERSDFTINLRVNLNFKLPGEYILVPSFTEPNKDGDCIRVSEKKADY 60
 Db 452 RDFELGRPSCTSSYIMNLREVSSVRLPQGLVIVPSTFEPNKGDCIRVSEKKAKA 511

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 12:48:39 ; Search time 11.1016 Seconds
(without alignments)
956.622 Million cell updates/sec

Title: US-09-884-319a-6
Perfect score: 1294
Sequence: 1 KNFPLNRARSDFFINLR.....PENTGTELDLISMLCFSVL 251

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Dackfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1294	100.0	251	1	US-08-726-525-6 Sequence 6, Appl
2	1294	100.0	251	2	US-08-487-942-6 Sequence 6, Appl
3	1294	100.0	251	2	US-08-726-036A-6 Sequence 6, Appl
4	1294	100.0	251	4	US-09-083-516-6 Sequence 6, Appl
5	1287	99.5	700	1	US-08-726-525-7 Sequence 7, Appl
6	1287	99.5	700	2	US-08-487-942-7 Sequence 7, Appl
7	1287	99.5	700	2	US-08-726-036A-7 Sequence 7, Appl
8	1287	99.5	700	3	US-09-422-869-23 Sequence 23, Appl
9	1287	99.5	700	4	US-09-083-516-7 Sequence 7, Appl
10	1213	93.7	700	4	US-09-308-345A-46 Sequence 46, Appl
11	762.5	58.9	714	4	US-09-308-345A-47 Sequence 47, Appl
12	757.5	58.5	714	4	US-09-422-869-22 Sequence 22, Appl
13	668.5	51.7	703	2	US-08-835-099A-1 Sequence 1, Appl
14	668.5	51.7	703	3	US-09-157-348-1 Sequence 1, Appl
15	668.5	51.7	712	2	US-08-835-099A-2 Sequence 2, Appl
16	668.5	51.7	712	2	US-08-835-099A-2 Sequence 2, Appl
17	646.5	50.0	739	4	US-09-653-839-6 Sequence 6, Appl
18	646.5	50.0	739	4	US-09-653-839-2 Sequence 2, Appl
19	633.5	49.0	703	3	US-09-422-869-27 Sequence 27, Appl
20	623	48.1	757	4	US-09-622-880B-1 Sequence 1, Appl
21	613	47.4	821	4	US-09-422-869-24 Sequence 24, Appl
22	611.5	47.3	803	4	US-09-308-345A-49 Sequence 49, Appl
23	609	46.6	821	4	US-09-622-880B-15 Sequence 15, Appl
24	597.5	46.2	821	4	US-09-308-345A-48 Sequence 48, Appl
25	593.5	45.9	866	4	US-09-422-869-28 Sequence 28, Appl
26	593.5	45.9	866	4	US-09-653-839-8 Sequence 8, Appl
27	593.5	45.9	723	4	US-09-653-839-4 Sequence 4, Appl

28	465	35.9	266	3	US-09-032-523-7 Sequence 7, Appl
29	457	35.3	268	2	US-08-835-099A-9 Sequence 9, Appl
30	457	35.3	268	2	US-09-157-349-9 Sequence 9, Appl
31	430	33.2	248	3	US-09-032-523-1 Sequence 1, Appl
32	421	32.5	828	4	US-09-308-345A-19 Sequence 19, Appl
33	185.5	14.3	214	1	US-08-766-605-1 Sequence 1, Appl
34	185.5	14.3	214	2	US-09-094-212-1 Sequence 1, Appl
35	170	13.1	191	1	US-08-766-605-3 Sequence 3, Appl
36	170	13.1	191	2	US-09-094-212-3 Sequence 3, Appl
37	122	9.4	640	4	US-09-308-345A-2 Sequence 2, Appl
38	119	9.2	513	4	US-09-308-345A-4 Sequence 4, Appl
39	111.5	8.6	639	3	US-09-422-869-25 Sequence 25, Appl
40	111.5	8.6	640	4	US-09-308-345A-7 Sequence 7, Appl
41	111	8.6	456	1	US-08-464-164-2 Sequence 2, Appl
42	111	8.6	456	1	US-08-338-057-2 Sequence 2, Appl
43	111	8.6	456	2	US-08-668-416-2 Sequence 2, Appl
44	109	8.4	408	1	US-07-951-715A-21 Sequence 21, Appl
45	109	8.4	408	2	US-08-459-448A-21 Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-726-525-6
Sequence 6, Application US/08726525
Patent No. 5789181
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
NUMBER OF INVENTIONS: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,525
FILING DATE: 07-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,942
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI5258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 251 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-726-525-6
Query Match 100.0%; Score 1294; DB 1; Length 251;
Best Local Similarity 100.0%; Pred. No. 2.2e-135;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KNFPLNRARSDFFINLRKLPGEYILVSTFEPRNKDGDCIRVFSKKADY 60

|||||
Db 1 KNFELTNARSRSTFIMLRVNLNFKLPGEYILVSTFEPNKGDFCIRVSEKKADY 60
QY 61 QAVDEITANLEEDDISDDIDGRRRLFAQLAGDAISAFELQTIIRVLAKRQD1KS 120
Db 61 QAVDEITANLEEDDISDDIDGRRRLFAQLAGDAISAFELQTIIRVLAKRQD1KS 120
QY 121 DGFSTETCKINWMDLSDSGSKLGKEFYILMTKIQOKYRIEIDVDRSGTMSYEMRK 180
Db 121 DGFSTETCKINWMDLSDSGSKLGKEFYILMTKIQOKYRIEIDVDRSGTMSYEMRK 180
QY 181 ALEAGFMPQOLHOVIYARFADQLIIDFNFVRCVLETLTKIFKQDPENTGTIEL 240
Db 181 ALEAGFMPQOLHOVIYARFADQLIIDFNFVRCVLETLTKIFKQDPENTGTIEL 240
QY 241 DLISMLCSYVL 251
Db 241 DLISMLCSYVL 251

RESULT 2

US-08-487-942-6
Sequence 6, Application US/08487942

Patent No. 5817476
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
TITLE OF INVENTION: BINDING
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,942
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 251 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-942-6

Query Match 100.0%; Score 1294; DB 2; Length 251;

Best Local Similarity 100.0%; Pred. No. 2,2e-135;

Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNFELTNARSRSTFIMLRVNLNFKLPGEYILVSTFEPNKGDFCIRVSEKKADY 60
Db 1 KNFELTNARSRSTFIMLRVNLNFKLPGEYILVSTFEPNKGDFCIRVSEKKADY 60
QY 61 QAVDEITANLEEDDISDDIDGRRRLFAQLAGDAISAFELQTIIRVLAKRQD1KS 120
Db 61 QAVDEITANLEEDDISDDIDGRRRLFAQLAGDAISAFELQTIIRVLAKRQD1KS 120

|||||
QY 121 DGFSTETCKINWMDLSDSGSKLGKEFYILMTKIQOKYRIEIDVDRSGTMSYEMRK 180
Db 121 DGFSTETCKINWMDLSDSGSKLGKEFYILMTKIQOKYRIEIDVDRSGTMSYEMRK 180
QY 181 ALEAGFMPQOLHOVIYARFADQLIIDFNFVRCVLETLTKIFKQDPENTGTIEL 240
Db 181 ALEAGFMPQOLHOVIYARFADQLIIDFNFVRCVLETLTKIFKQDPENTGTIEL 240
QY 241 DLISMLCSYVL 251
Db 241 DLISMLCSYVL 251

RESULT 3

US-08-726-036A-6
Sequence 6, Application US/08726036A

Patent No. 5981482
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
TITLE OF INVENTION: BINDING
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,036A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 251 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-726-036A-6

Query Match 100.0%; Score 1294; DB 2; Length 251;

Best Local Similarity 100.0%; Pred. No. 2,2e-135;

Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNFELTNARSRSTFIMLRVNLNFKLPGEYILVSTFEPNKGDFCIRVSEKKADY 60
Db 1 KNFELTNARSRSTFIMLRVNLNFKLPGEYILVSTFEPNKGDFCIRVSEKKADY 60
QY 61 QAVDEITANLEEDDISDDIDGRRRLFAQLAGDAISAFELQTIIRVLAKRQD1KS 120
Db 61 QAVDEITANLEEDDISDDIDGRRRLFAQLAGDAISAFELQTIIRVLAKRQD1KS 120
QY 121 DGFSTETCKINWMDLSDSGSKLGKEFYILMTKIQOKYRIEIDVDRSGTMSYEMRK 180
Db 121 DGFSTETCKINWMDLSDSGSKLGKEFYILMTKIQOKYRIEIDVDRSGTMSYEMRK 180
QY 181 ALEAGFMPQOLHOVIYARFADQLIIDFNFVRCVLETLTKIFKQDPENTGTIEL 240

Db 181 ALEAGFKPCQJHVIYARFADQDLIDPNDVRCVLETLFKIKQDPENTGTIEL 240
QY 241 DLISWLCFSYL 251
Db 241 DLISWLCFSYL 251

RESULT 4
US-09-083-516-6
; Sequence 6, Application US/09083516
; Patent No. 630086
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; TITLE OF INVENTION: BINDING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,516
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,942
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 251 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-083-516-6
Query Match 100.0%; Score 1294; DB 4; Length 251;
Best Local Similarity 100.0%; Pred. No. 2.2e-135;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 DLISWLCFSYL 251
Db 241 DLISWLCFSYL 251

RESULT 5
US-08-726-525-7
; Sequence 7, Application US/08726525
; Patent No. 5789181
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
; TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
; TITLE OF INVENTION: BINDING
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,525
; FILING DATE: 07-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,942
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-726-525-7
Query Match 99.5%; Score 1287; DB 1; Length 700;
Best Local Similarity 99.6%; Pred. No. 6.3e-134;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 690 DLISWLCFSYL 700

RESULT 6

US-08-487-942-7
; Sequence 7, Application US/08487942
; Patent No. 5817476

GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,942
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-487-942-7

Query Match 99.5%; Score 1287; DB 2; Length 700;
Best Local Similarity 99.6%; Pred. No. 6.3e-134;

Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNEFLTNRRARERSDPTFINLREVNLNFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 60
Db 450 KNEFLTNRRARERSDPTFINLREVNLNFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 509
QY 61 QAVDEIEANLEEFIDSEDDIDGVRRLFAQLAGDAEISAFELQTLIRVLAARODIKS 120
Db 510 QAVDEIEANLEEFIDSEDDIDGVRRLFAQLAGDAEISAFELQTLIRVLAARODIKS 569
QY 121 DGFSLIETCKIMVMDLSDSGSKLGKEFYILMTKIOKYOKIYREIDVDRSGTMSYEMRK 180
Db 570 DGFSLIETCKIMVMDLSDSGSKLGKEFYILMTKIOKYOKIYREIDVDRSGTMSYEMRK 629
QY 181 ALEBAGFKMPCQLHOVIYARFADQLIIDFNVFVCLVRLTLEKIFKQDLPENTGTIEL 240
Db 630 ALEBAGFKMPCQLHOVIYARFADQLIIDFNVFVCLVRLTLEKIFKQDLPENTGTIEL 689
QY 241 DLISWLCFSYL 251
Db 690 DLISWLCFSYL 700

RESULT 7

US-08-726-036A-7
; Sequence 7, Application US/08726036A
; Patent No. 5981482

GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,036A
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-726-036A-7

Query Match 99.5%; Score 1287; DB 2; Length 700;
Best Local Similarity 99.6%; Pred. No. 6.3e-134;

Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNEFLTNRRARERSDPTFINLREVNLNFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 60
Db 450 KNEFLTNRRARERSDPTFINLREVNLNFKLPGEYILVPSFEPNKGDCICIRVSEKKADY 509
QY 61 QAVDEIEANLEEFIDSEDDIDGVRRLFAQLAGDAEISAFELQTLIRVLAARODIKS 120
Db 510 QAVDEIEANLEEFIDSEDDIDGVRRLFAQLAGDAEISAFELQTLIRVLAARODIKS 569
QY 121 DGFSLIETCKIMVMDLSDSGSKLGKEFYILMTKIOKYOKIYREIDVDRSGTMSYEMRK 180
Db 570 DGFSLIETCKIMVMDLSDSGSKLGKEFYILMTKIOKYOKIYREIDVDRSGTMSYEMRK 629
QY 181 ALEBAGFKMPCQLHOVIYARFADQLIIDFNVFVCLVRLTLEKIFKQDLPENTGTIEL 240
Db 630 ALEBAGFKMPCQLHOVIYARFADQLIIDFNVFVCLVRLTLEKIFKQDLPENTGTIEL 689
QY 241 DLISWLCFSYL 251
Db 690 DLISWLCFSYL 700

RESULT 8

US-09-422-869-23
; Sequence 23, Application US/09422869
; Patent No. 6235481

GENERAL INFORMATION:
APPLICANT: POLONSKY, KENNETH S.
APPLICANT: HORIKAWA, YUKIO
APPLICANT: ODA, NAOHISA
APPLICANT: COX, NANCY J.
APPLICANT: SREENAN, SEAMUS
APPLICANT: ZHOU, YUN-PING
APPLICANT: OTANI, KENICHI
APPLICANT: HANIS, CRAIG L.
APPLICANT: BELL, GRAEME I.
TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
FILE REFERENCE: ARCD:307
CURRENT APPLICATION NUMBER: US/09/422,869
CURRENT FILING DATE: 1999-10-21
EARLIER APPLICATION NUMBER: 60/134,175
EARLIER FILING DATE: 1999-05-13
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 23
LENGTH: 700
TYPE: PRT
ORGANISM: Human
US-09-422-869-23

Query Match 99.5%; Score 1287; DB 3; Length 700;
Best Local Similarity 99.6%; Pred. No. 6.3e-134;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFFLTNRARSDFTFNLREVNLNFKLPGEYILVPSFEPNKGDFCIRVSEKKADY 60
DB 450 KNFFLTNRARSDFTFNLREVNLNFKLPGEYILVPSFEPNKGDFCIRVSEKKADY 509
QY 61 QAVDEIEANLEEFISDDIDDGRRFLFAQLAGDAEISAFELQTLIRVLAARODIKS 120
DB 510 QAVDEIEANLEEFISDDIDDGRRFLFAQLAGDAEISAFELQTLIRVLAARODIKS 569
QY 121 DGFSTETCKIMVMDSDSGKLGKEFYILMTKIQYOKIYREIDVDRSGTMNSYEMRK 180
DB 570 DGFSTETCKIMVMDSDSGKLGKEFYILMTKIQYOKIYREIDVDRSGTMNSYEMRK 629
QY 181 ALEAGFKMPCQLHOVIYARFADDDLLIDFNFVCLVRLTFLKIFKQDLPENTGTEL 240
DB 630 ALEAGFKMPCQLHOVIYARFADDDLLIDFNFVCLVRLTFLKIFKQDLPENTGTEL 689
QY 241 DLISWLCFSYL 251
DB 690 DLISWLCFSYL 700

RESULT 9
US-09-083-516-7
Sequence 7, Application US/09083516
Patent No. 6300086
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
TITLE OF INVENTION: BINDING
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
City: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/083,516
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,942
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-083-516-7

Query Match 99.5%; Score 1287; DB 4; Length 700;
Best Local Similarity 99.6%; Pred. No. 6.3e-134;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFFLTNRARSDFTFNLREVNLNFKLPGEYILVPSFEPNKGDFCIRVSEKKADY 60
DB 450 KNFFLTNRARSDFTFNLREVNLNFKLPGEYILVPSFEPNKGDFCIRVSEKKADY 509
QY 61 QAVDEIEANLEEFISDDIDDGRRFLFAQLAGDAEISAFELQTLIRVLAARODIKS 120
DB 510 QAVDEIEANLEEFISDDIDDGRRFLFAQLAGDAEISAFELQTLIRVLAARODIKS 569
QY 121 DGFSTETCKIMVMDSDSGKLGKEFYILMTKIQYOKIYREIDVDRSGTMNSYEMRK 180
DB 570 DGFSTETCKIMVMDSDSGKLGKEFYILMTKIQYOKIYREIDVDRSGTMNSYEMRK 629
QY 181 ALEAGFKMPCQLHOVIYARFADDDLLIDFNFVCLVRLTFLKIFKQDLPENTGTEL 240
DB 630 ALEAGFKMPCQLHOVIYARFADDDLLIDFNFVCLVRLTFLKIFKQDLPENTGTEL 689
QY 241 DLISWLCFSYL 251
DB 690 DLISWLCFSYL 700

RESULT 10
US-09-308-345A-46
Sequence 46, Application US/09308345A
Patent No. 6569665
GENERAL INFORMATION:
APPLICANT: BOEHM, Thomas;
APPLICANT: DEAR, Neil T.
TITLE OF INVENTION: No. 6569665el calpains, their preparation and use
FILE REFERENCE: 0050/47576
CURRENT APPLICATION NUMBER: US/09/308,345A
CURRENT FILING DATE: 1999-05-19
NUMBER OF SEQ ID NOS: 49
SOFTWARE: WordPerfect v. 6.1
SEQ ID NO 46
LENGTH: 700
TYPE: PRT
ORGANISM: mouse
US-09-308-345A-46

Query Match 93.7%; Score 1213; DB 4; Length 700;
Best Local Similarity 93.2%; Pred. No. 1e-125;
Matches 234; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 1 KNFFLTNRARSDFTFNLREVNLNFKLPGEYILVPSFEPNKGDFCIRVSEKKADY 60
DB 450 KNFFLTNRARSDFTFNLREVNLNFKLPGEYILVPSFEPNKGDFCIRVSEKKADY 509

```

Db      450  KNEFLTRARERSDTFIMLREVNLNRFKLPGEYIVLPSTFEPHKGDPEDIVFSEKKADY 509
QY      61  QAVDEIEANLEEFDISDDIDGFRRLFAQAGEDAEISAFELQTLIRVLAARODIKS 120
        |||:||||:| | :|:||||| | | | | | | | | | | | | | | | | | | | | | | |
Db      510  QAVDEIEANLEEFIDANEDIDGFRRLFVOLAGEDAEISAFELQTLIRVLAARODIKS 569
QY      121  DGSFIETCKIMVMDLSDGSGKLGKFEYILMTKIQYOKIYREIDVDRSGTMSYEMR 180
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      570  DGSFIETCKIMVMDLSDGSGKLGKFEYILMTKIQYOKIYREIDVDRSGTMSYEMR 629
QY      181  ALEAGFEMPCQLHQVIVARFADQOLIIDFNFRCLVRLTETLTKIKOLDPENTGTIEL 240
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      630  ALEAGFEMPCQLHQVIVARFADQOLIIDFNFRCLVRLTETLTKIKOLDPENTGTIEL 689
QY      241  DLJISMLCFSVL 251
        :| | | | |
Db      690  NLASMLSEFSVL 700

```

```

RESULT 11
US-09-308-345A-47
; Sequence 47, Application US/09308345A
; Patent No. 6569665
; GENERAL INFORMATION:
; APPLICANT: BOEHM, Thomas;
; APPLICANT: DEAR, Neil T.
; TITLE OF INVENTION: No. 6569665el calpains, their preparation and use
; FILE REFERENCE: 0050/47576
; CURRENT APPLICATION NUMBER: US/09/308,345A
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: WordPerfect v. 6.1
; SEQ ID NO 47
; LENGTH: 714
; TYPE: PRT
; ORGANISM: human
US-09-308-345A-47

```

```

Query Match      58.9%; Score 762.5; DB 4; Length 714;
Best Local Similarity 55.9%; Pred. No. 9.8e-76;
Matches 138; Conservative 57; Mismatches 51; Indels 1; Gaps 1;

```

```

QY      1  KNEFLTRARERSDTFIMLREVNLNRFKLPGEYIVLPSTFEPHKGDPEDIVFSEKKADY 60
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      462  RDEFLANASRASEGFIMLREVSTRFRLPGEYIVVSTFEPHKGDPEDIVFSEKSGT 521
QY      61  QAVDEIEANL-EEFDISEDDIDGFRRLFAQAGEDAEISAFELQTLIRVLAARODIK 119
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      522  VELDQIOANLPDEOVLSEEEIDENFKALFQAGEDMEISVKEFLRTILNRIISKRDLR 581
QY      120  SDGFIETCKIMVMDLSDGSGKLGKFEYILMTKIQYOKIYREIDVDRSGTMSYEMR 179
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      582  TKGFSIESCRSVNIMMDRGNGKLGVEFNILMNIRNYLSIFKFPDLDSGMSAYEMR 641
QY      180  KALEAGFEMPCQLHQVIVARFADQOLIIDFNFRCLVRLTETLTKIKOLDPENTGTIE 239
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      642  MAIESAGFLNKKLYELITRSEPDIAVDPNFVCCIVRLTETLTKIKOLDPENTGTIE 701
QY      240  LDJISML 246
        || | |
Db      702  FDLFKWL 708

```

```

RESULT 12
US-09-422-869-22
; Sequence 22, Application US/09422869
; Patent No. 6235461
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREEMAN, SEAMOS

```

```

; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 22
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Human
US-09-422-869-22

```

```

Query Match      58.5%; Score 757.5; DB 3; Length 714;
Best Local Similarity 55.5%; Pred. No. 3.5e-75;
Matches 137; Conservative 57; Mismatches 52; Indels 1; Gaps 1;

```

```

QY      1  KNEFLTRARERSDTFIMLREVNLNRFKLPGEYIVLPSTFEPHKGDPEDIVFSEKKADY 60
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      462  RDEFLANASRASEGFIMLREVSTRFRLPGEYIVVSTFEPHKGDPEDIVFSEKSGT 521
QY      61  QAVDEIEANL-EEFDISEDDIDGFRRLFAQAGEDAEISAFELQTLIRVLAARODIK 119
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      522  VELDQIOANLPDEOVLSEEEIDENFKALFQAGEDMEISVKEFLRTILNRIISKRDLR 581
QY      120  SDGFIETCKIMVMDLSDGSGKLGKFEYILMTKIQYOKIYREIDVDRSGTMSYEMR 179
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      582  TKGFSIESCRSVNIMMDRGNGKLGVEFNILMNIRNYLSIFKFPDLDSGMSAYEMR 641
QY      180  KALEAGFEMPCQLHQVIVARFADQOLIIDFNFRCLVRLTETLTKIKOLDPENTGTIE 239
        |||:||||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      642  MAIESAGFLNKKLYELITRSEPDIAVDPNFVCCIVRLTETLTKIKOLDPENTGTIE 701
QY      240  LDJISML 246
        || | |
Db      702  FDLFKWL 708

```

```

RESULT 13
US-08-835-099A-1
; Sequence 1, Application US/08835099A
; Patent No. 5874277
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuori
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/835,099A
; FILING DATE: 04-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 083649/1996
; FILING DATE: 05-APR-1996

```

RESULT 14
 US-09-157-349-1
 : Sequence 1, Application US/09157349
 : Patent No. 6068990
 : GENERAL INFORMATION:
 : APPLICANT: SHINTANI, Yasushi
 : APPLICANT: NISHI, Kazuo
 : APPLICANT: KARAMOTO, Tomohito
 : TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 : TITLE OF INVENTION: AND USE
 : NUMBER OF SEQUENCES: 18
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
 : STREET: 130 Water Street
 : CITY: Boston
 : STATE: MA
 : COUNTRY: USA
 : ZIP: 02109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FASTSEQ for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/157,349
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/835,099

RESULT 15
 US-08-835-099A-2
 : Sequence 2, Application US/08835099A
 : Patent No. 5874277
 : GENERAL INFORMATION:
 : APPLICANT: SHINTANI, Yasushi
 : APPLICANT: NISHI, Kazuo
 : APPLICANT: KAWAMOTO, Tomohiro
 : TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
 : TITLE OF INVENTION: AND USE
 : NUMBER OF SEQUENCES: 18
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: DIKE, BROWNSTEIN, ROBERTS & CUSMAN, LLP
 : STREET: 130 Water Street
 : City: Boston
 : STATE: MA
 : COUNTRY: USA
 : ZIP: 02109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FASTSEQ for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/835,099A
 : FILING DATE: 04 APR-1997
 : CLASSIFICATION: 514
 : PRIOR APPLICATION DATA:

PT - used to identify inhibitors of the proteins for treatment of
PT inflammation

PS Claim 4; Page 35-36; 54pp; English.

CC AAW19991 represents a protein that has interleukin-1 receptor (IL-1-R)
CC intracellular ligand activity. IL-1-R intracellular ligand proteins
CC are used to screen for agents (e.g. antibodies) that are capable of
CC inhibiting or blocking the binding of an IL-1-R intracellular ligand
CC to the intracellular domain of IL-1-R, i.e. inhibitors of IL-1
CC activity. Such agents can be used to treat inflammatory conditions.

XX Sequence 251 AA;

Query Match 100.0%; Score 1294; DB 18; Length 251;

Best Local Similarity 100.0%; Pred. No. 1,1e-122;

Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINLRVYLNRFKLPPEGYILVPSFTEPNKGDGCIRVSEKKADY 60
DB 1 KNFFLTNRARERSDFINLRVYLNRFKLPPEGYILVPSFTEPNKGDGCIRVSEKKADY 60

QY 61 QAVDEIEANLEEFISDDIDGFRRLFAQAGDAEISAFELQTLIRVLAKRQDIKS 120

DB 61 QAVDEIEANLEEFISDDIDGFRRLFAQAGDAEISAFELQTLIRVLAKRQDIKS 120

QY 121 DGFSEITCKIMVMDLSDSGSKLGKEFYILMTKIQKQKIRREIDVDRSGTMNSYEMRK 180

DB 121 DGFSEITCKIMVMDLSDSGSKLGKEFYILMTKIQKQKIRREIDVDRSGTMNSYEMRK 180

QY 181 ALEEGFPMPCQLHGVYIARFADQLIIDFNFVRCVLRLETLFKIFKQDPENGTIEL 240

DB 181 ALEEGFPMPCQLHGVYIARFADQLIIDFNFVRCVLRLETLFKIFKQDPENGTIEL 240

QY 241 DLISWLCFSVL 251

DB 241 DLISWLCFSVL 251

RESULT 2

AAB37796 standard; Protein; 251 AA.

AC AAB37796;

DT 23-FEB-2001 (first entry)

DE Human interleukin-1 receptor intracellular ligand protein #3.

XX Human; interleukin-1; IL-1; IL-1alpha; IL-1beta; IL-1 receptor;
XX antiinflammatory; haemostatic; antibacterial; immunosuppressive;
XX immunomodulator; cardiast; cytostatic; neuroprotective; respiratory;
XX inflammation; infection; sepsis; cachexia; autoimmune disorder;
XX cardiovascular disorder; chronic myelogenous leukaemia;
XX multiple sclerosis; inflammatory bowel disease; Crohn's disease.

XX Homo sapiens.

OS WO20064479-A1.

PN 02-NOV-2000.

PD 26-APR-2000; 2000WO-US11700.

PF 27-APR-1999; 99US-0301274.

PR (ANTI-) ANTIBODY SYSTEMS INC.

PI Fredeking TM, Ignatyev GM;

PR WPI; 2000-679646/66.

XX Novel compositions comprising tetracycline or tetracycline-like

PT compounds for the treatment and/or prevention of acute inflammatory
PT responses and diseases, e.g. septic shock and immune complex-induced
PT colitis -

PS Disclosure; Page 158-159; 183pp; English.

CC The present sequence is given in a specification relating to novel
CC compositions and methods containing tetracycline or tetracycline-like
CC compounds for treating and/or preventing acute inflammatory responses and
CC diseases. Such diseases include acute inflammatory conditions associated
CC with viral haemorrhagic diseases (including diseases caused by
CC Bunyaviridae, Flaviviridae or Arenaviridae viruses),
CC parasitic diseases, bacterial infections, sepsis, cachexia, autoimmune
CC disorders, acute cardiovascular events, chronic myelogenous leukaemia and
CC transplanted bone marrow-induced graft-versus-host disease, septic shock,
CC immune complex-induced colitis, cerebrospinal fluid inflammation,
CC multiple sclerosis, inflammatory responses associated with trauma,
CC systemic inflammatory response syndrome (SIRS), adult respiratory
CC distress syndrome (ARDS), acute liver failure, inflammatory bowel disease
CC and Crohn's disease.

XX Sequence 251 AA;

Query Match 100.0%; Score 1294; DB 21; Length 251;

Best Local Similarity 100.0%; Pred. No. 1,1e-122;

Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINLRVYLNRFKLPPEGYILVPSFTEPNKGDGCIRVSEKKADY 60

DB 1 KNFFLTNRARERSDFINLRVYLNRFKLPPEGYILVPSFTEPNKGDGCIRVSEKKADY 60

QY 61 QAVDEIEANLEEFISDDIDGFRRLFAQAGDAEISAFELQTLIRVLAKRQDIKS 120

DB 61 QAVDEIEANLEEFISDDIDGFRRLFAQAGDAEISAFELQTLIRVLAKRQDIKS 120

QY 121 DGFSEITCKIMVMDLSDSGSKLGKEFYILMTKIQKQKIRREIDVDRSGTMNSYEMRK 180

DB 121 DGFSEITCKIMVMDLSDSGSKLGKEFYILMTKIQKQKIRREIDVDRSGTMNSYEMRK 180

QY 181 ALEEGFPMPCQLHGVYIARFADQLIIDFNFVRCVLRLETLFKIFKQDPENGTIEL 240

DB 181 ALEEGFPMPCQLHGVYIARFADQLIIDFNFVRCVLRLETLFKIFKQDPENGTIEL 240

QY 241 DLISWLCFSVL 251

DB 241 DLISWLCFSVL 251

RESULT 3

AAW19992 standard; Protein; 700 AA.

AC AAW19992;

DT 27-AUG-1997 (first entry)

DE Human CAMP used to identify inhibitors of interleukin-1 activity.

XX IL; interleukin; receptor; ligand; screening assay; inhibitor;
XX IL-1 mediated response; inflammation; inflammatory; antibody;
XX intracellular domain; CAMP; calcium activated neutral protease.

XX Homo sapiens.

OS WO9640907-A1.

PN 19-DEC-1996.

PF 06-MAY-1996; 96WO-US06363.

PR 07-JUN-1995; 95US-0487942.

XX (GENET) GENETICS INST INC.

XX Graham J, Lin L;
PI
XX WPI; 1997-052315/05.
DR
XX Interleukin-1 receptor intracellular ligand proteins and related DNA
PT - used to identify inhibitors of the proteins for treatment of
PT inflammation
XX
PS Claim 14; Page 36-38; 54pp; English.
XX
CC AAW19992 represents human calcium activated neutral protease (CANP).
CC This protein was found to have an area of high homology with an
CC interleukin-1 receptor (IL-1-R) intracellular ligand (encoded by cDNA
CC clone 14w, see AAT71218) and thus will display some of the same
CC properties of this protein. IL-1-R intracellular ligand proteins are
CC used to screen for agents (e.g. antibodies) that are capable of
CC inhibiting or blocking the binding of an IL-1-R intracellular ligand
CC to the intracellular domain of IL-1-R, i.e. inhibitors of IL-1
CC activity. Such agents can be used to treat inflammatory conditions.
CC
XX
SQ Sequence 700 AA;
Query Match 99.5%; Score 1287; DB 18; Length 700;
Best Local Similarity 99.6%; Pred. No. 2.3e-121;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KNFELTNARERSDPTFNLREVLRNFKLPGEYILVPSFEPNKGDCIRVFSKKADY 60
DB 450 KNFELTNARERSDPTFNLREVLRNFKLPGEYILVPSFEPNKGDCIRVFSKKADY 509
QY 61 QAVDEIFANLEEFIDISEDIDGFRRLFAQLAGDAEISAFELOTILIRVYAKKODIKS 120
DB 510 QAVDEIFANLEEFIDISEDIDGFRRLFAQLAGDAEISAFELOTILIRVYAKKODIKS 569
QY 121 DGFSEITCKINWMDLSDSGSKLGKEFYIIMTKIQYOKIYREIDVDRSGTMNSYEMRK 180
DB 570 DGFSEITCKINWMDLSDSGSKLGKEFYIIMTKIQYOKIYREIDVDRSGTMNSYEMRK 629
QY 181 ALEAGFPMPCQLHGVYARFADQLIIDFNFVNCIVLRLETFKIFQDPENGTIEL 240
DB 630 ALEAGFPMPCQLHGVYARFADQLIIDFNFVNCIVLRLETFKIFQDPENGTIEL 689
QY 241 DLISWLCFSVL 251
DB 690 DLISWLCFSVL 700
RESULT 4
AAB37797
ID AAB37797 standard; Protein; 700 AA.
XX
AC AAB37797;
XX
DT 23-FEB-2001 (first entry)
XX
DE Human interleukin-1 receptor intracellular ligand protein #4.
XX
KW Human; interleukin-1; IL-1; IL-1alpha; IL-1beta; IL-1 receptor;
KW antiinflammatory; haemostatic; antibacterial; immunosuppressive;
KW immunomodulator; cardiac; cytosolic; neuroprotective; respiratory;
KW inflammation; infection; sepsis; cachexia; autoimmune disorder;
KW cardiovascular disorder; chronic myelogenous leukaemia;
KW multiple sclerosis; inflammatory bowel disease; Crohn's disease.
XX
OS Homo sapiens.
XX
PN WO200064479-A1.
XX
PD 02-NOV-2000.
XX
PF 26-APR-2000; 2000WO-US11700.
XX

PR 27-APR-1999; 99US-0301274.
XX
XX (ANTI-) ANTIBODY SYSTEMS INC.
XX
XX Fredeking TM, Ignatyev GM;
XX
DR WPI; 2000-679646/66.
XX
XX
PT Novel compositions comprising tetracycline or tetracycline-like
PT compounds for the treatment and/or prevention of acute inflammatory
PT responses and diseases, e.g. septic shock and immune complex-induced
PT colitis -
XX
PS Disclosure; Page 159-162; 183pp; English.
XX
CC The present sequence is given in a specification relating to novel
CC compositions and methods containing tetracycline or tetracycline-like
CC compounds for treating and/or preventing acute inflammatory responses and
CC diseases. Such diseases include acute inflammatory conditions associated
CC with viral haemorrhagic diseases (including diseases caused by
CC Bunyaviridae, Flaviviridae, or Arenaviridae viruses),
CC parasitic diseases, bacterial infections, sepsis, cachexia, autoimmune
CC disorders, acute cardiovascular events, chronic myelogenous leukaemia and
CC transplanted bone marrow-induced graft-versus-host disease, septic shock,
CC immune complex-induced colitis, cerebrospinal fluid inflammation,
CC multiple sclerosis, inflammatory responses associated with trauma,
CC systemic inflammatory response syndrome (SIRS), adult respiratory
CC distress syndrome (ARDS), acute liver failure, inflammatory bowel disease
CC and Crohn's disease.
CC
XX
SQ Sequence 700 AA;
Query Match 99.5%; Score 1287; DB 21; Length 700;
Best Local Similarity 99.6%; Pred. No. 2.3e-121;
Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KNFELTNARERSDPTFNLREVLRNFKLPGEYILVPSFEPNKGDCIRVFSKKADY 60
DB 450 KNFELTNARERSDPTFNLREVLRNFKLPGEYILVPSFEPNKGDCIRVFSKKADY 509
QY 61 QAVDEIFANLEEFIDISEDIDGFRRLFAQLAGDAEISAFELOTILIRVYAKKODIKS 120
DB 510 QAVDEIFANLEEFIDISEDIDGFRRLFAQLAGDAEISAFELOTILIRVYAKKODIKS 569
QY 121 DGFSEITCKINWMDLSDSGSKLGKEFYIIMTKIQYOKIYREIDVDRSGTMNSYEMRK 180
DB 570 DGFSEITCKINWMDLSDSGSKLGKEFYIIMTKIQYOKIYREIDVDRSGTMNSYEMRK 629
QY 181 ALEAGFPMPCQLHGVYARFADQLIIDFNFVNCIVLRLETFKIFQDPENGTIEL 240
DB 630 ALEAGFPMPCQLHGVYARFADQLIIDFNFVNCIVLRLETFKIFQDPENGTIEL 689
QY 241 DLISWLCFSVL 251
DB 690 DLISWLCFSVL 700
RESULT 5
AAB86128
ID AAB86128 standard; Protein; 700 AA.
XX
AC AAB86128;
XX
DT 27-JUL-2001 (first entry)
XX
DE Human calpain 80kDa subdomain protein fragment.
XX
KW Calpain; calcium-activated cysteine proteinase; human; spatial structure;
KW Ca-activated cysteine proteinase; protein coordinate data; treatment;
KW structure-function study; ischemic condition; muscular dystrophy; tumor;
KW muscular; antitumor.
XX
OS Homo sapiens.
XX

XX EPI108779-A2.
 XX 20-JUN-2001.
 XX 13-DEC-2000; 2000EP-0127369.
 XX 14-DEC-1999; 99DE-1060225.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX (PROT-) PROTEROS BIOSTRUCTURES GMBH.
 XX Strobl S, Fernandez-Catalan C, Bode W, Huber R, Suzuki K;
 XX WPI; 2001-376928/40.
 XX Spatial structures containing calpain-derived polypeptides, useful for
 XX identifying calpain modulators and substrates, potentially useful e.g.
 XX as antitumor agents -
 XX Claim 15; Fig 4; 182pp; German.
 XX This invention describes the novel spatial structure of human and rat
 XX neutral calcium-activated cysteine protease (calpain) family. The spatial
 XX structure (especially crystalline forms) are used for structure-function
 XX studies, particularly for identifying (pseudo)substrates, inhibitors and
 XX activators of calpains, potentially useful for treatment of ischemic
 XX conditions, muscular dystrophy and/or tumors. The products of the
 XX invention have anti-ischemic, muscular and antitumor activity. This
 XX sequence represents the human calpain 80kDa subunit described in the
 XX method of the invention.

SQ Sequence 700 AA:

Query Match 99.5%; Score 1287; DB 22; Length 700;
 Best Local Similarity 99.6%; Pred. No. 2.3e-121;
 Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINLREVLNRFKLPGEYILVPTFEPNKGDFCIRVSEKKADY 60
 DB 450 KNFFLTNRARERSDFINLREVLNRFKLPGEYILVPTFEPNKGDFCIRVSEKKADY 509
 QY 61 QAVDEIFANLEEFISDDIDGGRRLFAQLAGDAISAFELOTILRRVLAKRQDIKS 120
 DB 510 QAVDEIFANLEEFISDDIDGGRRLFAQLAGDAISAFELOTILRRVLAKRQDIKS 569
 QY 121 DGFSETCKIWMMDSDSGSKGLGKEFYILMTKIQKQKIYREIDVDRSGTMSYEMRK 180
 DB 570 DGFSETCKIWMMDSDSGSKGLGKEFYILMTKIQKQKIYREIDVDRSGTMSYEMRK 629
 QY 181 ALEAGFMPQCOLHOVIYARFADDLIIDFNFVRCVLRLLETKIFKQOLDPENTGTIEL 240
 DB 630 ALEAGFMPQCOLHOVIYARFADDLIIDFNFVRCVLRLLETKIFKQOLDPENTGTIEL 689
 QY 241 DLISWLCFSVL 251
 DB 690 DLISWLCFSVL 700

RESULT 6
 AAE25059

ID AAE25059 standard; Protein; 700 AA.

XX AAE25059;

DT 30-OCT-2002 (first entry)

XX Human calpain protein #2.

XX Human; calpain; nervous system disorder; amyotrophic lateral sclerosis;
 XX Parkinson's disease; dementia; genito-urinary system disorder; stroke;
 XX Alzheimer's disease; multiple sclerosis; benign prostatic hyperplasia;
 XX urinary incontinence; gene therapy; cytoskeletal; neuropathic;

KW neuroprotective.
 XX Homo sapiens.
 XX WO200248326-A2.
 XX 20-JUN-2002.
 XX 14-DEC-2001; 2001WO-EP14819.
 XX 14-DEC-2000; 2000US-255058P.
 XX (FARB) BAYER AG.
 XX Ramakrishnan S;
 XX WPI; 2002-537625/57.
 XX New human calpain polypeptide, useful for treating peripheral and
 XX central nervous system disorder and genito-urinary system disorders
 XX including urinary incontinence and benign prostatic hyperplasia -
 XX Disclosure; Page 101-104; 110pp; English.

SQ Sequence 700 AA:

Query Match 99.5%; Score 1287; DB 23; Length 700;
 Best Local Similarity 99.6%; Pred. No. 2.3e-121;
 Matches 250; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNFFLTNRARERSDFINLREVLNRFKLPGEYILVPTFEPNKGDFCIRVSEKKADY 60
 DB 450 KNFFLTNRARERSDFINLREVLNRFKLPGEYILVPTFEPNKGDFCIRVSEKKADY 509
 QY 61 QAVDEIFANLEEFISDDIDGGRRLFAQLAGDAISAFELOTILRRVLAKRQDIKS 120
 DB 510 QAVDEIFANLEEFISDDIDGGRRLFAQLAGDAISAFELOTILRRVLAKRQDIKS 569
 QY 121 DGFSETCKIWMMDSDSGSKGLGKEFYILMTKIQKQKIYREIDVDRSGTMSYEMRK 180
 DB 570 DGFSETCKIWMMDSDSGSKGLGKEFYILMTKIQKQKIYREIDVDRSGTMSYEMRK 629
 QY 181 ALEAGFMPQCOLHOVIYARFADDLIIDFNFVRCVLRLLETKIFKQOLDPENTGTIEL 240
 DB 630 ALEAGFMPQCOLHOVIYARFADDLIIDFNFVRCVLRLLETKIFKQOLDPENTGTIEL 689
 QY 241 DLISWLCFSVL 251
 DB 690 DLISWLCFSVL 700

RESULT 7
 AAB86130

ID AAB86130 standard; Protein; 700 AA.

XX AAB86130;

DT 27-JUL-2001 (first entry)

XX Rat calpain 80kDa subdomain protein fragment.

XX Calpain; calcium-activated cysteine proteinase; human; spatial structure;
 XX Ca-activated cysteine proteinase; protein coordinate data; treatment;
 XX structure-function study; ischemic condition; muscular dystrophy; tumor;

KM muscular; antitumor.
 XX
 OS Rattus norvegicus.
 XX
 PM EPI108779-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 13-DEC-2000; 2000EP-0127369.
 XX
 PR 14-DEC-1999; 99DE-1060225.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX (PROT-) PROTEROS BIOSTRUCTURES GMBH.
 XX
 PI Strobl S, Fernandez-Catalan C, Bode W, Huber R, Suzuki K;
 XX
 DR WPI; 2001-376928/40.
 XX
 PT Spatial structures containing calpain-derived polypeptides, useful for
 PT identifying calpain modulators and substrates, potentially useful e.g.
 PT as antitumor agents
 XX
 PS Claim 15; Fig 6; 182pp; German.
 XX
 CC This invention describes the novel spatial structure of human and rat
 CC neutral calcium-activated cysteine protease (calpain) family. The spatial
 CC structure (especially crystalline forms) are used for structure-function
 CC studies, particularly for identifying (pseudo)substrates, inhibitors and
 CC activators of calpains, potentially useful for treatment of ischemic
 CC conditions, muscular dystrophy and/or tumors. The products of the
 CC invention have anti-ischemic, muscular and antitumor activity. This
 CC sequence represents the rat calpain 80kDa subunit described in the
 CC method of the invention.
 CC
 SQ Sequence 700 AA:
 Query Match 93.5%; Score 1210; DB 22; Length 700;
 Best Local Similarity 92.4%; Pred. No. 1.4e-113;
 Matches 232; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

QY 1 KNFELTNARSRSTFIMLRVNLNFKLPGEYIIVSTFEPNKGDFCIRVSEKKADY 60
 DB 450 KNFELTNARSRSTFIMLRVNLNFKLPGEYIIVSTFEPNKGDFCIRVSEKKADY 509
 QY 61 QAVDELEANLEEFDISDDIDGFRRLFAOLAGDAEISAFELQTLIRVLAKEKODIKS 120
 DB 510 QVDELEANLEEFDISDDIDGFRRLFAOLAGDAEISAFELQTLIRVLAKEKODIKS 569
 QY 121 DGFSIETCKIIVMDLSDSGSKLGKFEFYILMTKIQKYOKIYREIDVDSGMTNSYEMRK 180
 DB 570 DGFSIETCKIIVMDLSDSGSKLGKFEFYILMTKIQKYOKIYREIDVDSGMTNSYEMRK 629
 QY 181 ALAEAGFMPQQLHOYIYARFADQLIIDFNFVACLVRLLETFIFKQDPENTGTEIQL 240
 DB 630 ALAEAGFMPQQLHOYIYARFADQLIIDFNFVACLVRLLETFIFKQDPENTGTEIQL 689
 QY 241 DLISWLCFSVL 251
 DB 690 DLISWLCFSVL 700

RESULT 8
 ABG70047
 ID ABG70047 standard: Protein; 206 AA.
 AC ABG70047;
 XX
 DT 21-OCT-2002 (first entry)
 XX
 DE Human prey protein for Shigella ospB #6.
 XX
 KM Prey protein; ospB; ospD1; ipad; ipac; ipaH9.8; ospG;

KM ospC1; Shigella; shigellosis; bacillary dysentery; antibacterial;
 KM Yeast two-hybrid system; protein-protein interaction; SID;
 KM selected interacting domain; human.
 XX
 OS Homo sapiens.
 XX
 PM WO200257303-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 11-JAN-2002; 2002WO-EP00777.
 XX
 PR 12-JAN-2001; 2001US-261130P.
 XX
 PA (HYBR-) HYBRIGENICS.
 XX
 PI Legrain P;
 XX
 DR WPI; 2002-599706/64.
 DR N-PSDB; ABS51440.
 XX
 PT New complex of protein-protein interactions between a bait Shigella
 PT flexneri polypeptide and a prey mammalian or human placenta polypeptide
 PT for treating or preventing bacillary dysentery in a mammal or human
 XX
 PS Claim 7; Page 60; 162pp; English.
 XX
 CC The invention relates to a complex of protein-protein interactions
 CC between a Shigella flexneri polypeptide (e.g. ospB, ospD1, ipad, ipac,
 CC ipaH9.8, ospG and ospC1) and a mammalian polypeptide defined in the
 CC specification. The complexes are formed using the yeast two-hybrid
 CC system. Also included are (1) a recombinant host cell expressing the
 CC interactions between the Shigella flexneri polypeptide and a mammalian
 CC polypeptide defined in the specification; (2) selecting a modulating
 CC compound that inhibits or activates the protein-protein interactions;
 CC (3) a modulating compound obtained from the method of (2); (4) a SID
 CC comprising the human polypeptides appearing as ABG70042-ABG70242;
 CC (5) a SID polynucleotide or its fragment or variant comprising
 CC encoding the above polypeptides a vector comprising (5);
 CC (6) a recombinant host cell containing the vector; and (10) a protein
 CC chip comprising Shigella flexneri polypeptide and a mammalian polypeptide
 CC defined in the specification. A pharmaceutical composition comprising the
 CC compound, polypeptide or polynucleotide is useful for treating or
 CC preventing shigellosis (bacillary dysentery) in a human or mammal.
 CC The present sequence represents a human prey protein isolated by the
 CC yeast two-hybrid assay, forming a complex of the invention with a
 CC shigella protein.
 CC
 SQ Sequence 206 AA:
 Query Match 81.2%; Score 1051; DB 23; Length 206;
 Best Local Similarity 99.5%; Pred. No. 3.6e-98;
 Matches 205; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 46 GDFCIRVSEKKADYQAVDELEANLEEFDISDDIDGFRRLFAOLAGDAEISAFELQ 105
 DB 1 GDFCIRVSEKKADYQAVDELEANLEEFDISDDIDGFRRLFAOLAGDAEISAFELQ 60
 QY 106 TILRRVLAKEKODIKSGFSIETCKIIVMDLSDSGSKLGKFEFYILMTKIQKYOKIYREI 165
 DB 61 TILRRVLAKEKODIKSGFSIETCKIIVMDLSDSGSKLGKFEFYILMTKIQKYOKIYREI 120
 QY 166 DVDRSGTNSYEMRKALAEAGFKMPQQLHOYIYARFADQLIIDFNFVACLVRLLETFIFK 225
 DB 121 DVDRSGTNSYEMRKALAEAGFKMPQQLHOYIYARFADQLIIDFNFVACLVRLLETFIFK 180
 QY 226 IFKQDPENTGTEIQLISWLCFSVL 251
 DB 181 IFKQDPENTGTEIQLISWLCFSVL 206

RESULT 9

AAB51707
ID AAB51707 standard; Protein: 191 AA.
AC AAB51707;
XX
XX 26-FEB-2001 (first entry)
DT
XX
XX Gene 36 human secreted protein homologous amino acid sequence #147.
DE
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KM cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KM fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KM hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KM cerebrovascular disorder; nervous system disorder; ocular disorder;
KM wound healing; skin aging; food additive; preservative.
XX
XX Homo sapiens.
XX
XX WO200061620-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09069.
XX
XX 09-APR-1999; 99US-0128702.
XX
XX 20-JAN-2000; 2000US-0177049.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI: 2000-619225/59.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure: Page 521; 540pp; English.
XX
XX The polynucleotide sequences given in AAC93364 to AAC93412 encode the
XX human secreted proteins given in AAB51620 to AAB51668. AAB51669 to
XX AAB51722 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissues
XX and cells the genes are expressed in. Examples of activities include:
XX antirheumatic; immunosuppressive; antirheumatic; antiproliferative;
XX cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
XX neuroprotective; antibacterial; virucide; fungicide; and
XX ophthalmological. The polynucleotides and polypeptides can be used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX in diagnosing a pathological condition or susceptibility to a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases, hyperproliferative disorders, cardiovascular
XX disorders, cerebrovascular disorders, angiogenesis, nervous system
XX disorders, infections caused by bacteria, viruses and fungi and ocular
XX disorders. The polypeptides can also be used to aid wound healing and
XX epithelial cell proliferation, to prevent skin aging due to sunburn, to
XX maintain organs before transplantation, for supporting cell culture of
XX primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities. AAC93365 to AAC93363 and
XX AAB51619 represent sequences which are used in the exemplification of the
XX present invention.
XX
XX Sequence 191 AA:
XX
XX Query Match 74.9%; Score 969; DB 21; Length 191;
XX Best Local Similarity 99.5%; Pred. NO. 6.5e-90;
XX Matches 190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 61 QAVDDEIANLEEFISDDIDDGRRRLFAQLAGDAEISAFELQTLIRRYLAKRQDIKS 120

DB 1 QAVDDEIANLEEFISDDIDDGRRRLFAQLAGDAEISAFELQTLIRRYLAKRQDIKS 60
QY 121 DGFSTETCKIWMMDSDSGSKGLKEFYILMTKIOYKQRYREIDVRSCTMSYEMRK 180
DB 61 DGFSTETCKIWMMDSDSGSKGLKEFYILMTKIOYKQRYREIDVRSCTMSYEMRK 120
QY 181 ALEBAGEFMPQCLHGVYARFADDLIIFDNFVRCVLVRETLKIFKQDLPENTGTTEL 240
DB 121 ALEBAGEFMPQCLHGVYARFADDLIIFDNFVRCVLVRETLKIFKQDLPENTGTTEL 180
QY 241 DLISMLCFSVL 251
DB 181 DLISMLCFSVL 191
RESULT 10
AAB51708
ID AAB51708 standard; Protein: 191 AA.
AC AAB51708;
XX
XX 26-FEB-2001 (first entry)
XX
XX Human secreted protein sequence encoded by gene 36 SEQ ID NO:148.
XX
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KM cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KM fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KM hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KM cerebrovascular disorder; nervous system disorder; ocular disorder;
KM wound healing; skin aging; food additive; preservative.
XX
XX Homo sapiens.
XX
XX WO200061620-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US09069.
XX
XX 09-APR-1999; 99US-0128702.
XX
XX 20-JAN-2000; 2000US-0177049.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI: 2000-619225/59.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure: Page 522; 540pp; English.
XX
XX The polynucleotide sequences given in AAC93364 to AAC93412 encode the
XX human secreted proteins given in AAB51620 to AAB51668. AAB51669 to
XX AAB51722 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissues
XX and cells the genes are expressed in. Examples of activities include:
XX antirheumatic; immunosuppressive; antirheumatic; antiproliferative;
XX cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
XX neuroprotective; antibacterial; virucide; fungicide; and
XX ophthalmological. The polynucleotides and polypeptides can be used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX in diagnosing a pathological condition or susceptibility to a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases, hyperproliferative disorders, cardiovascular
XX disorders, cerebrovascular disorders, angiogenesis, nervous system

